2-2015

Shrinkage estimation of dynamic panel data models with interactive fixed effects

Xun LU
Hong Kong University of Science and Technology

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

SMU ECONOMICS & STATISTICS
WORKING PAPER SERIES

Shrinkage Estimation of Dynamic Panel Data Models with Interactive Fixed Effects

Xun Lu and Liangjun Su

February 2015

Paper No. 02-2015

Shrinkage Estimation of Dynamic Panel Data Models with Interactive Fixed Effects *

Xun Lu[a] and Liangjun Su[b]

[a] Department of Economics, HKUST

[b] School of Economics, Singapore Management University

February 2, 2015

Abstract

We consider the problem of determining the number of factors and selecting the proper regressors in linear dynamic panel data models with interactive fixed effects. Based on the preliminary estimates of the slope parameters and factors *a la* Bai and Ng (2009) and Moon and Weidner (2014a), we propose a method for simultaneous selection of regressors and factors and estimation through the method of adaptive group Lasso (*least absolute shrinkage and selection operator*). We show that with probability approaching one, our method can correctly select all relevant regressors and factors and shrink the coefficients of irrelevant regressors and redundant factors to zero. Further, we demonstrate that our shrinkage estimators of the nonzero slope parameters exhibit some oracle property. We conduct Monte Carlo simulations to demonstrate the superb finite-sample performance of the proposed method. We apply our method to study the determinants of economic growth and find that in addition to three common unobserved factors selected by our method, government consumption share has negative effects, whereas investment share and lagged economic growth have positive effects on economic growth.

JEL Classification: C13, C23, C51

Key Words: Adaptive Lasso; Dynamic panel; Factor selection; Group Lasso; Interactive fixed effects; Oracle property; Selection consistency

*The authors gratefully thank the Co-editor Cheng Hsiao, the Associate Editor, and three anonymous referees for their many constructive comments on the previous version of the paper. They express their sincere appreciation to Jushan Bai, Songnian Chen, Xiaohong Chen, Han Hong, Lung-Fei Lee, Qi Li, Hyungsik Roger Moon, Peter C. B. Phillips, and Peng Wang for discussions on the subject matter and valuable comments on the paper. They also thank the conference and seminar participants at the 3rd Shanghai Econometrics Workshop, Tsinghua International Conference in Econometrics, 2013, and Peking University for their valuable comments. Su gratefully acknowledges the Singapore Ministry of Education for Academic Research Fund under grant number MOE2012-T2-2-021. Address Correspondence to: Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903; E-mail: ljsu@smu.edu.sg, Phone: +65 6828 0386.

1 Introduction

We consider a panel data model with interactive fixed effects as proposed and studied in Pesaran (2006), Bai (2009), Moon and Weidner (2014a, 2014b), Pesaran and Tosetti (2011), Greenaway-McGrevy et al. (2012), Su and Jin (2012), Su et al. (2015), among others. This model has been widely applied in empirical research, as it allows more flexible modeling of heterogeneity than traditional fixed effects models and provides an effective way to model cross section dependence that is widely present in macro and financial data. To use this model, we need to determine the number of factors in the multi-factor error component and select the proper regressors to be included in the model. This paper provides a novel automated estimation method that combines both estimation of parameters of interest and selection of the number of factors and regressors.

Specifically, we consider the following interactive fixed-effects panel data model

$$Y_{it} = \beta^{0\prime} X_{it} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}, \; i = 1, ..., N, \; t = 1, \ldots, T, \tag{1.1}$$

where X_{it} is a $K_0 \times 1$ vector of regressors, β^0 is the corresponding vector of slope coefficients, λ_i^0 is an $R_0 \times 1$ vector of unknown factor loadings, F_t^0 is an $R_0 \times 1$ vector of unknown common factors, and ε_{it} is the idiosyncratic error term. Here the factor structure $\lambda_i^{0\prime} F_t^0$ is referred to as interactive fixed effects in Bai (2009) and Moon and Weidner (2014a, 2014b) as one allows both λ_i^0 and F_t^0 to be correlated with elements of X_{it} and $\lambda_i^{0\prime} F_t^0 + \varepsilon_{it}$ is called the multi-factor error structure in Pesaran (2006). We are interested in estimating β^0, λ_i^0 and F_t^0. It has been argued that the factor structure can capture more flexible heterogeneity across individuals and over time than the traditional fixed-effects model. The latter takes the form $Y_{it} = \beta^{0\prime} X_{it} + \alpha_i^0 + \xi_t^0 + \varepsilon_{it}$ and can be thought of as a special case of the interactive fixed-effects panel data model by letting $F_t^0 = (1, \xi_t^0)'$ and $\lambda_i^0 = (\alpha_i^0, 1)'$, where α_i^0 and ξ_t^0 are individual-specific and time-specific fixed effects, respectively. When X_{it} is absent in (1.1), the model becomes the pure factor model studied in Bai and Ng (2002) and Bai (2003), among others.

Given the correct number R_0 of factors and the proper regressors X_{it}, several estimation methods have been proposed in the literature. For example, Pesaran (2006) proposes common correlated effects (CCE) estimators; Bai (2009) and Moon and Weidner (2014a, 2014b) provide estimators based on Gaussian quasi-maximum likelihood estimation (QMLE) and the principal component analysis (PCA). To apply the latter methods, we must first determine the number of factors and appropriate regressors to be included in the model. Nevertheless, in practice, we do not have *a priori* knowledge about the true number of factors in almost all cases. Also there may be a large number of potential regressors, some of which may be irrelevant. Thus it is desirable to use a parsimonious model by choosing a subset of regressors. The common procedure is to perform some model selection in the first step and then conduct estimation based on the selected regressors and the chosen number of factors. To select regressors, a wide range of methods can be adopted. For example, one can apply the Bayesian information criterion (BIC) or some cross-validation methods. To determine the number of factors, one can apply the information criteria proposed in Bai and Ng (2002) or the testing procedure introduced in Onatski (2009, 2010), Kapetanios (2010), or Ahn and Horenstein (2013). Bai and Ng (2006, 2007) provide some empirical examples of the determination of number of factors in economic applications. Hallin and Liśka (2007) study the determination of the number of factors in general dynamic factor models.

In this paper, we explore a different approach. We use shrinkage techniques to combine the estimation with the selection of the number of factors and regressors in a single step. Following Bai (2009) or Moon and Weidner (2014a, 2014b), we can set a maximum number of factors (R, say) and obtain the preliminary estimates of the slope parameters and factors. Then we consider a penalized least squares (PLS) regression of Y_{it} on X_{it} and the estimated factors via the adaptive (group) Lasso. We include two penalty terms in the PLS, one for the selection of regressors in X_{it} via adaptive Lasso and the other for the selection of the exact number of factors via adaptive group Lasso. Despite the use of estimated factors that have slow convergence rates, we show that our new method can *consistently* determine the number of factors, *consistently* select all relevant regressors, and shrink the estimates of the coefficients of irrelevant regressors and redundant factors to zero *with probability approaching 1* (w.p.a.1). We also demonstrate the *oracle* property of our method. That is, our estimator of the non-zero regression coefficients is asymptotically equivalent to the least squares estimator based on the factor-augmented regression where both the true number of factors and the set of relevant regressors are *known*. The bias-corrected version of our shrinkage estimator of the non-zero regression coefficients is asymptotically equivalent to Moon and Weidner's (2014b) bias-corrected QML estimator in the case where all regressors are *relevant* (i.e., there is no selection of regressors). In the presence of *irrelevant* regressors, the variance-covariance matrix for our shrinkage estimator of the non-zero coefficients is smaller than that of Moon and Weidner's QML estimator. In addition, we emphasize that even though Moon and Weidner (2014a) show that the limiting distribution of the QML estimator is independent of the number of factors used in the estimation as long as the number of factors does not fall below the true number of factors, we find that in finite samples the inclusion of redundant factors can result in significant loss of efficiency (see Section 4.3 for detail). For this reason, it is very important to include the correct number of factors in the model especially when the cross section or time dimension is not very large. Our shrinkage method effectively selects all *relevant* regressors and factor estimates and get rid of irrelevant regressors or redundant factor estimates.

There is a large statistics literature on the shrinkage type of estimation methods. See, for example, Tibshirani (1996) for the origin of Lasso, Knight and Fu (2000) for the first systematic study of the asymptotic properties of Lasso-type estimators, and Fan and Li (2001) for SCAD (*smoothly clipped absolute deviation*) estimators. Zou (2006) establishes the oracle property of adaptive Lasso; Yuan and Lin (2006) propose the method of group Lasso; Wang and Leng (2008) and Wei and Huang (2010) study the properties of adaptive group Lasso; Huang et al. (2008) study Bridge estimators in sparse high dimensional regression models. Recently there have been an increasing number of applications of the shrinkage techniques in the econometrics literature. For example, Caner (2009) and Fan and Liao (2014) consider covariate selection in GMM estimation. Belloni et al. (2013) and García (2011) consider selection of instruments in the GMM framework. Liao (2013) provides a shrinkage GMM method for moment selection and Cheng and Liao (2015) consider the selection of valid and relevant moments via penalized GMM. Liao and Phillips (2015) apply adaptive shrinkage techniques to cointegrated systems. Kock (2013) considers Bridge estimators of *static* linear panel data models with random or fixed effects. Caner and Knight (2013) apply Bridge estimators to differentiate a unit root from a stationary alternative. Caner and Han (2014) propose a Bridge estimator for pure factor models and shows the selection consistency. Cheng et al. (2014) provide an adaptive group Lasso estimator for pure factor structures with possible structural breaks. This paper adds to the literature by applying the shrinkage idea to panel data models with factor structures and considering generated regressors.

The method proposed in this paper has a wide range of applications. For example, it can be used to estimate a structural panel model that allows a more flexible form of heterogeneity. A specific example is to study cross-country economic growth. Let Y_{it} be the economic growth for country i in period t and X_{it} be a large number of potential observable causes of economic growth, such as physical capital investment, consumption, population growth, government consumption, and lagged economic growth, among others. Economic growth may also be caused by many unobservable common factors F_t^0. It is of great interest to know which observable causes are important to determine economic growth and the number of common unobserved factors that affect all countries' economic growth. Our new method is directly applicable to this important economic question. Another example of application is to forecast asset returns, as factor models are often used to model asset returns. Specifically, let Y_{it} be the excess returns on asset i in period t and X_{it} be observable factors such as Fama-French factors (small market capitalization and book-to-market ratio), divided yields, dividend payout ratio and consumption gap, among others. The asset returns may also be affected by an unknown number of common unobserved factors. Our method automatically selects the important observable factors and unobservable common factors. Thus it provides a powerful tool to predict future asset returns.

The paper is organized as follows. Section 2 introduces our adaptive group Lasso estimators. Section 3 analyzes their asymptotic properties. In Section 4, we report the Monte Carlo simulation results for our method and compare it with the methods of Bai and Ng (2002), Onatski (2009, 2010), and Ahn and Horenstein (2013). In Section 5, we apply our method to study the determinants of economic growth in the framework of dynamic panel data models with interactive fixed effects, and find that in addition to three common unobserved factors selected by our method, government consumption share has negative effects, whereas investment share and lagged economic growth have positive effects on economic growth. Final remarks are contained in Section 6. The proofs of all theorems are delegated to Appendix B. Additional materials are provided in the online supplementary Appendices C-F.

NOTATION. For an $m \times n$ real matrix A, we denote its transpose as A', its Frobenius norm as $\|A\|$ ($\equiv [\mathrm{tr}(AA')]^{1/2}$), its spectral norm as $\|A\|_{\mathrm{sp}}$ ($\equiv \sqrt{\mu_1(A'A)}$) and its Moore-Penrose generalized inverse as A^+, where $\equiv$ means "is defined as" and $\mu_s(\cdot)$ denotes the sth largest eigenvalue of a real symmetric matrix by counting eigenvalues of multiplicity multiple times. Note that the two norms are equal when A is a vector. We will frequently use the submultiplicative property of these norms and the fact that $\|A\|_{\mathrm{sp}} \leq \|A\| \leq \|A\|_{\mathrm{sp}} \mathrm{rank}(A)^{1/2}$. We also use $\mu_{\max}(B)$ and $\mu_{\min}(B)$ to denote the largest and smallest eigenvalues of a symmetric matrix B, respectively. We use $B > 0$ to denote that B is positive definite. Let $P_A \equiv A(A'A)^+ A'$ and $M_A \equiv I_m - P_A$, where I_m denotes an $m \times m$ identity matrix. The operator $\xrightarrow{P}$ denotes convergence in probability, $\xrightarrow{d}$ convergence in distribution, and plim probability limit. We use $(N,T) \to \infty$ to denote that N and T pass to infinity jointly.

2 Penalized Estimation of Panel Data Models with Interactive Fixed Effects

In this section, we consider penalized least squares (PLS) estimation of panel data models with interactive fixed effects where the number of unobservable factors is unknown and some observable regressors may be irrelevant.

2.1 Panel Data Models with Interactive Fixed Effects

We assume that the true model (1.1) is unknown, in particular, R_0 and K_0 are unknown. With a little bit abuse of notation, we consider their empirical model

$$Y_{it} = \beta^{0\prime} X_{it} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it},\ i = 1, ..., N,\ t = 1, \ldots, T, \tag{2.1}$$

where X_{it} is a $K \times 1$ vector of regressors that may contain lagged dependent variables, $\beta^0 \equiv \left(\beta_1^0, ..., \beta_K^0\right)'$ is a $K \times 1$ vector of unknown slope coefficients, F_t^0 and λ_i^0 are $R \times 1$ vectors of factors and factor loadings, respectively, and ε_{it} is the idiosyncratic error term. Here $\left\{\lambda_i^0\right\}$ and $\left\{F_t^0\right\}$ may be correlated with $\{X_{it}\}$. We consider estimation and inference on β^0 when the true number of factors R_0 ($\leq R$) is unknown and some variables in X_{it} may be irrelevant, i.e., $K_0 \leq K$. In the sequel, we allow both K and K_0 to pass to infinity as $(N, T) \to \infty$ but assume that R is fixed to facilitate the asymptotic analysis.

To proceed, let $X_{it,k}$ denote the kth element of X_{it} for $k = 1, ..., K$. Define

$$\begin{aligned}
Y_i &\equiv (Y_{i1}, \ldots, Y_{iT})',\ X_i \equiv (X_{i1}, \ldots, X_{iT})',\ \varepsilon_i \equiv (\varepsilon_{i1}, \ldots, \varepsilon_{iT})', \\
F^0 &\equiv \left(F_1^0, \ldots, F_T^0\right)',\ \lambda^0 \equiv \left(\lambda_1^0, ..., \lambda_N^0\right)',\ X_{i,\cdot k} \equiv (X_{i1,k}, ..., X_{iT,k})', \\
\mathbf{Y} &\equiv (Y_1, ..., Y_N)',\ \mathbf{X}_k \equiv (X_{1,\cdot k}, ..., X_{N,\cdot k})',\ \text{and } \boldsymbol{\varepsilon} \equiv (\varepsilon_1, ..., \varepsilon_N)'.
\end{aligned}$$

Apparently $\mathbf{Y}$, $\mathbf{X}_k$, and $\boldsymbol{\varepsilon}$ are all $N \times T$ matrices. Then we can write the model (2.1) in matrix form

$$\mathbf{Y} = \sum_{k=1}^{K} \beta_k^0 \mathbf{X}_k + \lambda^0 F^{0\prime} + \boldsymbol{\varepsilon}. \tag{2.2}$$

Without loss of generality (Wlog), we assume that only the first K_0 elements of X_{it} have nonzero slope coefficients, and write $X_{it} = (X_{it(1)}', X_{it(2)}')'$, where $X_{it(1)}$ and $X_{it(2)}$ are $K_0 \times 1$ and $(K - K_0) \times 1$ vectors, respectively, and the true coefficients of $X_{it(1)}$ are nonzero while those of $X_{it(2)}$ are zero. Accordingly, we decompose β^0 as $\beta^0 = (\beta_{(1)}^{0\prime}, \beta_{(2)}^{0\prime})' = (\beta_{(1)}^{0\prime}, 0')'$.

2.2 QMLE of $\left(\beta^0, \lambda^0, F^0\right)$

Given R and all regressors, following Bai (2009) and Moon and Weidner (2014a, 2014b), we consider the Gaussian QMLE $(\tilde{\beta}, \tilde{\lambda}, \tilde{F})$ of $\left(\beta^0, \lambda^0, F^0\right)$ which is given by

$$\left(\tilde{\beta}, \tilde{\lambda}, \tilde{F}\right) = \underset{(\beta, \lambda, F)}{\arg\min}\ \mathcal{L}_{NT}^0 \left(\beta, \lambda, F\right), \tag{2.3}$$

where

$$\mathcal{L}_{NT}^{0}\left(\beta,\lambda,F\right)\equiv\frac{1}{NT}\mathrm{tr}\left[\left(\mathbf{Y}-\sum_{k=1}^{K}\beta_{k}\mathbf{X}_{k}-\lambda F'\right)'\left(\mathbf{Y}-\sum_{k=1}^{K}\beta_{k}\mathbf{X}_{k}-\lambda F'\right)\right],\tag{2.4}$$

$\beta\equiv(\beta_1,...,\beta_K)'$ is a $K\times1$ vector, $F\equiv(F_1,...,F_T)'$ is a $T\times R$ matrix, and $\lambda\equiv(\lambda_1,...,\lambda_N)'$ is an $N\times R$ matrix. One can first obtain the profile-likelihood estimate $\tilde{\beta}$ and then the estimate $(\tilde{\lambda},\tilde{F})$ via the PCA method under the identification restrictions: $F'F/T=I_R$ and $\lambda'\lambda$ is a diagonal matrix. Namely, $(\tilde{\lambda},\tilde{F})$ solves

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(Y_i-X_i\tilde{\beta}\right)\left(Y_i-X_i\tilde{\beta}\right)'\right]\tilde{F}=\tilde{F}V_{NT}\text{ and }\tilde{\lambda}=T^{-1}\left(\mathbf{Y}-\sum_{k=1}^{K}\tilde{\beta}_k\mathbf{X}_k\right)\tilde{F},\tag{2.5}$$

where V_{NT} is a diagonal matrix consisting of the R largest eigenvalues of the above matrix in the square bracket, arranged in descending order. Moon and Weidner (2014a) show that as long as $R\geq R_0$, the limiting distribution of the QMLE for β is independent of R, the number of unobserved factors used in the estimation. Throughout the paper, we assume that $R\geq R_0$ and use $\tilde{\beta}^c=(\tilde{\beta}_1^c,...,\tilde{\beta}_K^c)'$ to denote the bias-corrected version of $\tilde{\beta}$ based on the formula in Moon and Weidner (2014b) or our supplementary Appendix F. After obtaining $\tilde{\beta}^c$, we obtain the final estimate $(\tilde{\lambda},\tilde{F})$ via (2.5) with $\tilde{\beta}$ replaced by $\tilde{\beta}^c$.

2.3 Penalized Least Squares Estimation of $\left(\beta^0,\lambda^*\right)$

We first present our PLS estimators and then provide some motivations for them. Our PLS estimator $(\hat{\beta},\hat{\lambda})$ are obtained as follows.

- Estimate model (2.1) with R factors and all K regressors and obtain $(\tilde{\lambda},\tilde{F})$ and $\tilde{\beta}^c$ as discussed in Section 2.2.
- Let $\hat{\mathbf{Y}}=\mathbf{Y}-\sum_{k=1}^{K}\tilde{\beta}_k^c\mathbf{X}_k$, $\hat{F}=(NT)^{-1}\hat{\mathbf{Y}}'\hat{\mathbf{Y}}\tilde{F}$, and $\hat{\Sigma}_{\hat{F}}=T^{-1}\hat{F}'\hat{F}$. Compute the R eigenvalues of $\hat{\Sigma}_{\hat{F}}$ arranged in descending order and denote them as $\tau_1,...,\tau_R$.
- Minimize the following PLS criterion function

$$Q_{\gamma}\left(\beta,\lambda\right)=\mathcal{L}_{NT}\left(\beta,\lambda\right)+\gamma_{1NT}\sum_{k=1}^{K}\frac{1}{\left|\tilde{\beta}_k^c\right|^{\kappa_1}}\left|\beta_k\right|+\frac{\gamma_{2NT}}{\sqrt{N}}\sum_{r=1}^{R}\frac{1}{\tau_r^{\kappa_2}}\left\|\lambda_{\cdot r}\right\|,\tag{2.6}$$

  where $\mathcal{L}_{NT}\left(\beta,\lambda\right)=\mathcal{L}_{NT}^0(\beta,\lambda,\hat{F})$, $\lambda_{\cdot r}$ denotes the rth column of λ, $\gamma=\gamma_{NT}=(\gamma_{1NT},\gamma_{2NT})$ is a vector of tuning parameters, and $\kappa_1,\ \kappa_2>0$ are usually taken as either 1 or 2. Let $(\hat{\beta},\hat{\lambda})=(\hat{\beta}\left(\gamma\right),\hat{\lambda}\left(\gamma\right))$ denote the solution to the above minimization problem.

Note that (2.6) contains two penalty terms, γ_{1NT} for the regression coefficients β_k's and γ_{2NT} for the loading vectors $\lambda_{\cdot r}$'s. Noting that $N^{-1/2}\left\|\lambda_{\cdot r}\right\|=O_P(1)$ under our Assumption A.1(iii) in Section 3.1 which apparently rules out the case of weak factors studied by Onatski (2012), we divide the second penalty term γ_{2NT} by $\sqrt{N}$. Note that the objective function in (2.6) is convex in (β,λ) so that the global minimizer of Q_γ can be found easily for any given tuning parameter γ. We frequently suppress the dependence of $(\hat{\beta},\hat{\lambda})$ on γ as long as no confusion arises. Below we will propose a data-driven method to choose γ. Also, note that we have used estimated factors $\hat{F}_t$ in (2.6).

As a referee points out, the idea to use group Lasso for selection of the number of factors has been around for some time. For example, Hirose and Konishi (2012) derive a model selection criterion for selecting factors in a pure factor model but they do not provide asymptotic analysis. In contrast, we consider both variable and factor selections in dynamic panel data models and offer systematic asymptotic analysis.

Our procedure is motivated by the literature on *adaptive group Lasso* (see, Yuan and Lin (2006), Zou (2006), Huang et al. (2008)). Now we provide some details. R is usually different from R_0 and one cannot expect $\tilde{F}$ to be a consistent estimator of F^0 or a rotational version of F^0. Define $H = H_{NT} = \left(N^{-1}\lambda^{0\prime}\lambda^0\right)\left(T^{-1}F^{0\prime}\tilde{F}\right)$. We can follow Bai and Ng (2002) and show that under certain regularity conditions, $\frac{1}{T}\left\|\hat{F}-F^0H\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$ and $\left\|\hat{F}_t - H'F_t^0\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$ for each t, where $\hat{F}_t'$ denotes the tth row of $\hat{F}$ and $\delta_{NT} = \min(\sqrt{N}, \sqrt{T})$. In addition, we show in Appendix A that $H = H_{NT}$ converges in probability to a sparse matrix

$$H^0 = [H^0_{(1)},\ \mathbf{0}_{R_0\times(R-R_0)}],$$

where $H^0_{(1)}$ is an $R_0\times R_0$ full rank matrix and $\mathbf{0}_{a\times b}$ denotes an $a\times b$ matrix of zeros. As a result, $\lambda_i^* = H^+\lambda_i^0$ also exhibits a sparse structure asymptotically, i.e., the last $(R-R_0)$ elements of λ_i^* converge in probability to zero. Using the above definitions of H and λ_i^*, we can rewrite (2.1) as[1]

$$Y_{it} = \beta^{0\prime}X_{it} + \lambda_i^{0\prime}H^{+\prime}H'F_t^0 + \varepsilon_{it} = \beta^{0\prime}X_{it} + \lambda_i^{*\prime}H'F_t^0 + \varepsilon_{it}. \tag{2.7}$$

The sparse nature of λ_i^* (and β^0) suggests that we can apply an adaptive group Lasso procedure as introduced above. Further, we show in Appendix A that $\tau_1, ..., \tau_{R_0}$ converge in probability to some finite positive numbers whereas $\tau_{R_0+1}, ..., \tau_R$ converge to zero at $\sqrt{N}$-rate. This means that $\hat{\Sigma}_{\hat{F}}$ provides the information on the sparsity nature of λ_i^*. This motivates us to use $\frac{1}{\tau_r^{\kappa 2}}$ as a weight in the second penalty term in (2.6).

3 Asymptotic Properties

In this section we study the asymptotic properties of the proposed adaptive group Lasso estimator $(\hat{\beta}, \hat{\lambda})$.

3.1 Estimation Consistency

Let $\bar{W}_{NT} = \frac{1}{NT}\sum_{i=1}^N X_i'X_i$ and $W_{NT} = \frac{1}{NT}\sum_{i=1}^N \tilde{X}_i'M_{F^0}\tilde{X}_i$ where $\tilde{X}_i \equiv X_i - \frac{1}{N}\sum_{j=1}^N \lambda_i^{0\prime}\left(N^{-1}\lambda^{0\prime}\lambda^0\right)^{-1}\lambda_j^0 X_j$. Let C denote a generic finite positive constant that may vary across lines. We make the following assumptions.

Assumption A.1 (i) $\sqrt{NT/K}\left\|\tilde{\beta}^c - \beta^0\right\| = O_P(1)$ and $\sqrt{NT}\left|\tilde{\beta}_k^c - \beta_k^0\right| = O_P(1)$ for each $k = 1, ..., K$.

(ii) $E\left\|F_t^0\right\|^8 \le C$ and $T^{-1}F^{0\prime}F^0 \xrightarrow{P} \Sigma_{F^0} > 0$ for some $R_0\times R_0$ matrix Σ_{F^0} as $T\to\infty$.

(iii) $E\left\|\lambda_i^0\right\|^8 \le C$ and $N^{-1}\lambda^{0\prime}\lambda^0 \xrightarrow{P} \Sigma_{\lambda^0} > 0$ for some $R_0\times R_0$ matrix Σ_{λ^0} as $N\to\infty$.

(iv) For $k = 1, ..., K$, $(NT)^{-1}E\left\|\mathbf{X}_k\right\|^2 \le C$.

[1] Noting that the $R_0\times R$ matrix H is right invertible, by Proposition 6.1.5 in Bernstein (2005, p.225) we have $H^+ = H'\left(HH'\right)^{-1}$, which further implies that $HH^+ = I_{R_0}$.

(v) $\|\boldsymbol{\varepsilon}\|_{\text{sp}} = O_P(\max(\sqrt{N}, \sqrt{T}))$.

(vi) For $k = 1, ..., K$, $(NT)^{-1} E\left[\text{tr}\left(\mathbf{X}_k \boldsymbol{\varepsilon}'\right)\right]^2 \leq C$.

(vii) $(NT)^{-1} E\left\|\lambda^{0\prime} \boldsymbol{\varepsilon} F^0\right\|^2 \leq C$.

(viii) There are two nonstochastic $K \times K$ matrices $\bar{W}_0$ and W_0 such that $\left\|\bar{W}_{NT} - \bar{W}_0\right\|_{\text{sp}} = o_P(1)$ and $\|W_{NT} - W_0\|_{\text{sp}} = o_P(1)$, where $\mu_{\max}\left(\bar{W}_0\right)$ and $\mu_{\min}(W_0)$ are bounded away from infinity and zero, respectively.

Assumption A.2 (i) $E(\varepsilon_{it}) = 0$ and $E\left(\varepsilon_{it}^8\right) \leq C$.

(ii) Let $\sigma_{ij,ts} = E(\varepsilon_{it}\varepsilon_{js})$. $\max_{1 \leq t \leq T} N^{-1} \sum_{i=1}^N \sigma_{ii,tt} \leq C$, $N^{-1} \sum_{i=1}^N \sum_{j=1}^N \max_{1 \leq t \leq T} |\sigma_{ij,tt}| \leq C$, $T^{-1} \sum_{t=1}^T \sum_{s=1}^T \max_{1 \leq i \leq N} |\sigma_{ii,ts}| \leq C$, and $(NT)^{-1} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T |\sigma_{ij,ts}| \leq C$.

(iii) For every (t, s), $E\left|N^{-1/2} \sum_{i=1}^N \left[\varepsilon_{it}\varepsilon_{is} - E(\varepsilon_{it}\varepsilon_{is})\right]\right|^4 \leq C$.

Assumption A.3 (i) As $(N, T) \to \infty$, $T/N^2 \to 0$, $N/T^2 \to 0$, and $K^2/\min(N, T) \to 0$.

(ii) As $(N, T) \to \infty$, $(TK_0)^{1/2}\gamma_{1NT} \to 0$, and $T^{1/2}\gamma_{2NT} \to 0$.

A.1(i) is a high-level assumption. Primitive conditions can be found in Moon and Weidner (2014a, 2014b) which ensure the $\sqrt{NT}$-consistency of a bias-corrected preliminary estimate when K is fixed and $R = R_0$. In the supplementary Appendix F, we extend the analysis to allow diverging K and $R > R_0$. As a referee points out, one can relax this assumption to allow for a non-bias-corrected estimator of β, in which case A1(i) would become $\left\|\tilde{\beta} - \beta^0\right\| = O_P\left(K^{1/2}\delta_{NT}^{-2}\right)$ and $\left\|\tilde{\beta}_k - \beta_k^0\right\| = O_P\left(\delta_{NT}^{-2}\right)$ for each $k = 1, 2, ..., K$, and more bias terms need to be corrected for the shrinkage estimator $\hat{\beta}$ than here. A.1(ii)-(iv) impose standard moment conditions on F_t^0, λ_i^0, and X_{it}; see, e.g., Bai and Ng (2002) and Bai (2003, 2009). Note that Bai and Ng (2002) assume only the fourth moment for F_t^0 but require that λ_i^0 be uniformly bounded. Moon and Weidner (2014a) demonstrate that A.1(v) can be satisfied for various error processes. A.1(vi) requires weak exogeneity of the regressor $\mathbf{X}_k$. A.1(vii) can be satisfied under various primitive conditions and it implies that $\left\|\lambda^{0\prime}\boldsymbol{\varepsilon}F^0\right\| = O_P(N^{1/2}T^{1/2})$ by Chebyshev inequality, which further implies that $\left\|\lambda^{0\prime}\boldsymbol{\varepsilon}\right\| = O_P(N^{1/2}T^{1/2})$ and $\left\|\boldsymbol{\varepsilon}F^0\right\| = O_P(N^{1/2}T^{1/2})$ under Assumptions A.1(ii)-(iii) by standard matrix operations. A.1(viii) requires that the large dimensional matrices $\bar{W}_{NT}$ and W_{NT} be well behaved asymptotically.

A.2 is adopted from Bai and Ng (2002) and Bai (2009). It allows for weak forms of both cross sectional dependence and serial dependence in the error processes. The first two parts of A.3(i) require that T should not grow too fast in comparison with N and vice versa; the last part of A.3(i), namely, $K^2/\min(N, T) \to 0$, is needed to ensure that the estimation of the $K \times 1$ vector β plays asymptotic negligible role on the estimation of the factors and factor loadings. A.3(ii) is a condition that ensures a preliminary $\sqrt{T}$-rate of consistency of our shrinkage estimator $\hat{\beta}$ (see Theorem 3.1 below) and it essentially says that the two penalty terms cannot be too large.

The following theorem establishes the consistency of the shrinkage estimator $(\hat{\beta}, \hat{\lambda})$.

Theorem 3.1 *Suppose Assumptions A.1, A.2, and A.3(i)-(ii) hold. Then*

$$\left\|\hat{\beta} - \beta^0\right\| = O_P\left(T^{-1/2}\right), \text{ and } N^{-1}\left\|\hat{\lambda}H' - \lambda^0\right\|^2 = \frac{1}{N}\sum_{i=1}^N \left\|H\hat{\lambda}_i - \lambda_i^0\right\|^2 = O_P\left(T^{-1}\right).$$

Remark 1. Theorem 3.1 establishes the *preliminary* $\sqrt{T}$-rate of consistency for $\hat{\beta}$ (in Euclidean norm) and the usual T-rate of consistency for the cross sectional average of the squared deviations between the estimated factor loadings (with rotation) and the true factor loadings. The former is a preliminary rate and will be improved later on. The latter is the best rate of consistency one can obtain. It is worth mentioning that the second part of the result in the above theorem in general does not imply $\frac{1}{N}\left\|\hat{\lambda}-\lambda^0 H^{+\prime}\right\|^2 = O_P\left(T^{-1}\right)$ unless $R = R_0$. To see why, notice that

$$\begin{aligned}\frac{1}{N}\left\|\hat{\lambda}-\lambda^0 H^{+\prime}\right\|^2 &= \frac{1}{N}\left\|\hat{\lambda}\left(I_R - H'H^{+\prime}\right)+\left(\hat{\lambda}H'-\lambda^0\right)H^{+\prime}\right\|^2 \\ &\geq \frac{1}{N}\left\|\hat{\lambda}\left(I_R - H'H^{+\prime}\right)\right\|^2 - \frac{1}{N}\left\|\left(\hat{\lambda}H'-\lambda^0\right)H^{+\prime}\right\|^2.\end{aligned}$$

Even though Theorem 3.1 implies that the second term is $O_P\left(T^{-1}\right)$, the first term does not vanish asymptotically as $H'H^{+\prime} \neq I_R$ for any $R > R_0$. Nevertheless, by the triangle inequality, the fact that H has full row rank asymptotically, and Assumption A.1(iii), we can readily show that $N^{-1}||\hat{\lambda}||^2 = O_P(1)$.

3.2 Selection Consistency

To study the selection consistency, we write $\hat{\beta} = (\hat{\beta}_{(1)}', \hat{\beta}_{(2)}')'$ and $\hat{\lambda} = (\hat{\lambda}_{(1)}, \hat{\lambda}_{(2)})$, where $\hat{\beta}_{(1)}$ and $\hat{\beta}_{(2)}$ are column vectors of dimensions K_0 and $K-K_0$, respectively, and $\hat{\lambda}_{(1)}$ and $\hat{\lambda}_{(2)}$ are $N \times R_0$ and $N \times (R - R_0)$ matrices, respectively.

To state the next result, we augment Assumption A.3 with one further condition.

Assumption A.3 (iii) As $(N,T) \to \infty$, $(NT)^{\kappa_1/2}T^{1/2}\gamma_{1NT} \to \infty$, and $N^{\kappa_2/2}T^{1/2}\gamma_{2NT} \to \infty$.

Clearly, A.3(iii) requires that the two penalty terms should not be too small. The next theorem establishes the selection consistency of our adaptive group Lasso procedure.

Theorem 3.2 *Suppose Assumptions A.1, A.2, and A.3(i)-(iii) hold. Then*

$$P\left(\left\|\hat{\beta}_{(2)}\right\| = 0 \text{ and } \left\|\hat{\lambda}_{(2)}\right\| = 0\right) \to 1 \text{ as } (N,T) \to \infty.$$

Remark 2. Theorem 3.2 says that with w.p.a.1 all the zero elements in β^0 and all the factor loadings of the redundant factor estimates must be estimated to be exactly zero. On the other hand, by Theorem 3.1, we know that the estimates of the nonzero elements in β^0 and the factor loadings of the non-redundant factor estimates must be consistent. This implies that w.p.a.1, all the relevant regressors and estimated factors must be identified by nonzero coefficients and nonzero factor loadings, respectively. Put together, Theorems 3.1 and 3.2 imply that the adaptive group Lasso has the ability to identify the true regression model with the correct number of factors consistently.

3.3 Oracle Property

Decompose $\tilde{F} = (\tilde{F}_{(1)}, \tilde{F}_{(2)})$ where $\tilde{F}_{(1)}$ and $\tilde{F}_{(2)}$ are $T \times R_0$ and $T \times (R - R_0)$ submatrices, respectively. Analogously, let $\hat{F} = (\hat{F}_{(1)}, \hat{F}_{(2)})$ and $H = (H_{(1)}, H_{(2)})$, where $\hat{F}_{(l)} = (NT)^{-1}\hat{\mathbf{Y}}'\hat{\mathbf{Y}}\tilde{F}_{(l)}$ and $H_{(l)} = H_{(l)NT} = \left(N^{-1}\lambda^{0\prime}\lambda^0\right)\left(T^{-1}F^{0\prime}\tilde{F}_{(l)}\right)$ for $l = 1, 2$. Let $F_{(1)}^* \equiv F^0 H_{(1)}$ and $\lambda_{(1)}^* \equiv \lambda^0 H_{(1)}^{+\prime}$. Let $F_{t(1)}^{*\prime}$ and $\lambda_{i(1)}^{*\prime}$ denote the tth and ith rows of $F_{(1)}^*$ and $\lambda_{(1)}^*$, respectively. Define $\hat{F}_{t(1)}$ and $\hat{\lambda}_{i(1)}$ analogously. Further, write

$X_i = (X_{i(1)}, X_{i(2)})$ where $X_{i(1)}$ and $X_{i(2)}$ are $T \times K_0$ and $T \times (K - K_0)$ submatrices of X_i, respectively. Let $\hat{D}_{F^0} \equiv (NT)^{-1} \sum_{i=1}^{N} X'_{i(1)} M_{F^0} X_{i(1)}$ and $C'_{NT} \equiv (NT)^{-2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_j^{0\prime} F^{0\prime} \tilde{F}_{(1)} \lambda^*_{i(1)} X'_{i(1)} M_{F^0} X_j$. Let $\mathcal{D} \equiv \sigma(F^0, \lambda^0)$, the sigma-field generated by (F^0, λ^0), and $E_{\mathcal{D}}(A) \equiv E(A|\mathcal{D})$. Define

$$\begin{aligned}
\mathbb{B}_{1NT} &\equiv (NT)^{-5/2} \sum_{i=1}^{N} X'_{i(1)} F^*_{(1)} \left(F^{*\prime}_{(1)} F^*_{(1)}\right)^{-1} \tilde{F}'_{(1)} F^0 \lambda^{0\prime} \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \lambda^0 F^{0\prime} \tilde{F}_{(1)} \lambda^*_{i(1)}, \\
\mathbb{B}_{2NT} &\equiv (NT)^{-3/2} \sum_{i=1}^{N} X'_{i(1)} M_{F^0} \boldsymbol{\varepsilon}' \boldsymbol{\varepsilon} \tilde{F}_{(1)} \lambda^*_{i(1)}, \\
\mathbb{B}_{3NT} &\equiv (NT)^{-3/2} \sum_{i=1}^{N} X'_{i(1)} F^*_{(1)} \left(F^{*\prime}_{(1)} F^*_{(1)}\right)^{-1} \tilde{F}'_{(1)} F^0 \lambda^0 \boldsymbol{\varepsilon}' \varepsilon_i, \text{ and} \\
\mathbb{B}_{4NT} &\equiv (NT)^{-1/2} \sum_{i=1}^{N} \left[X_{i(1)} - E_{\mathcal{D}}\left(X_{i(1)}\right)\right]' P_{F^0} \varepsilon_i.
\end{aligned}$$

To study the oracle property of $\hat{\beta}_{(1)}$ and $\hat{\lambda}_{i(1)}$, we add the following assumptions.

Assumption A.4 (i) There exists a $K_0 \times K_0$ matrix $D_{F^0} > 0$ such that $\left\|\hat{D}_{F^0} - D_{F^0}\right\|_{\text{sp}} = o_P(1)$.

(ii) There exists a $K \times K_0$ matrix C_0 such that $\|C_{NT} - C_0\|_{\text{sp}} = o_P(1)$.

(iii) $\max_{1 \le k \le K} E\left\|F^{0\prime} \boldsymbol{\varepsilon}' \mathbf{X}_k \boldsymbol{\varepsilon}' \lambda^0\right\|^2 = O\left(N^2 T^2 (N+T)\right)$.

(iv) $\max_{1 \le i \le N} E\left\|\lambda^{0\prime} \boldsymbol{\varepsilon} \varepsilon_i\right\|^2 = O(T(N+T))$.

Assumption A.5 (i) Let $\bar{Z}_i \equiv X_{i(1)} - P_{F^0} E_{\mathcal{D}}(X_{i(1)}) - M_{F^0} \bar{\mathcal{X}}_{i1NT} + \left[X_i - E_{\mathcal{D}}(X_i) - M_{F^0} \bar{\mathcal{X}}_{i2NT}\right] W_0^{-1} C_0$, where $\bar{\mathcal{X}}_{i1NT} = \frac{1}{N} \sum_{j=1}^{N} \lambda_i^{0\prime} (T^{-1} F^{0\prime} \tilde{F}_{(1)}) \lambda^*_{j(1)} E_{\mathcal{D}}(X_{j(1)})$ and $\bar{\mathcal{X}}_{i2NT} = \frac{1}{N} \sum_{j=1}^{N} \lambda_i^{0\prime} \left(N^{-1} \lambda^{0\prime} \lambda^0\right)^{-1} \lambda_j^0 E_{\mathcal{D}}(X_j)$. There exists Θ_{NT} such that $\frac{1}{\sqrt{NT}} \mathbb{C}_{K_0} \sum_{i=1}^{N} \bar{Z}'_i \varepsilon_i \xrightarrow{d} N\left(0, \lim_{(N,T)\to\infty} \mathbb{C}_{K_0} \Theta_{NT} \mathbb{C}'_{K_0}\right)$ for any $\ell \times K_0$ nonrandom matrix $\mathbb{C}_{K_0}$ such that $\mathbb{C}_{K_0} \mathbb{C}'_{K_0} \to \mathbb{C}$, where $\ell \in [1, K_0]$ is a fixed finite integer and Θ_{NT} has eigenvalues that are bounded away from zero and infinity for sufficiently large (N, T).

(ii) There exists $\Theta_{i,F^*_{(1)}} > 0$ such that $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} F^*_{t(1)} \varepsilon_{it} \xrightarrow{d} N(0, \Theta_{i,F^*_{(1)}})$.

Assumption A.6 (i) As $(N, T) \to \infty$, $K_0^{1/2}(T^{1/2} N^{-1} + N^{1/2} T^{-1}) \to 0$.

(ii) As $(N, T) \to \infty$, $(NTK_0)^{1/2} \gamma_{1NT} \to 0$ and $(NT)^{1/2} \gamma_{2NT} \to 0$.

Assumptions A.4(i)-(ii) are weak as one can readily show that both $\hat{D}_{F^0}$ and C_{NT} are $O_P(1)$ in the case of fixed K. The positive definiteness of D_{F^0} is ensured by Assumption A.1(viii). C_0 is generally not zero in A.4(ii), but it can be zero under fairly restrictive conditions on the data generating processes for $\left\{X_{it}, \varepsilon_{it}, F_t^0, \lambda_i^0\right\}$. See Greenaway-McGrevy et al. (2012, GHS hereafter) and the discussion in Remark 4 below. A.4(iii)-(iv) are high level assumptions. A.5 parallels Assumption E in Bai (2009) which is also a high level assumption. Note that both cross sectional and serial dependence and heteroskedasticity are allowed in the error terms. We verify these assumptions in the supplementary appendix by allowing lagged dependent variables in X_{it}. A.6 is needed to obtain the oracle property for our adaptive group Lasso estimator.

The following theorem establishes the asymptotic distributions of both $\hat{\beta}_{(1)}$ and $\hat{\lambda}_{i(1)}$.

Theorem 3.3 *Suppose Assumptions A.1-A.6 hold. Let* $\mathbb{V}_{NT} = D_{F^0}^{-1} \Theta_{NT} D_{F^0}^{-1}$ *and* $\mathbb{C}_{K_0}$ *be as defined in Assumption A.5(i). Then* *(i)* $\mathbb{C}_{K_0}\left[\sqrt{NT}\left(\hat{\beta}_{(1)}-\beta^0_{(1)}\right)-\mathbb{B}_{NT}\right]\xrightarrow{D} N\left(0,\lim_{(N,T)\to\infty}\mathbb{C}_{K_0}\mathbb{V}_{NT}\mathbb{C}'_{K_0}\right),$

(ii) $\sqrt{T}\left(\hat{\lambda}_{i(1)}-H^{+}_{(1)}\lambda^0_i\right)\xrightarrow{D} N\left(0,\Sigma^{-1}_{F^*_{(1)}}\Theta_{i,F^*_{(1)}}\Sigma^{-1}_{F^*_{(1)}}\right),$

where $\mathbb{B}_{NT}=D^{-1}_{\hat{F}_{(1)}}\left(\mathbb{B}_{1NT}-\mathbb{B}_{2NT}-\mathbb{B}_{3NT}-\mathbb{B}_{4NT}\right),\ \Sigma_{F^*_{(1)}}=H^{0\prime}_{(1)}\Sigma_{F^0}H^0_{(1)},$ *and* $H^0_{(1)}$ *is the probability limit of* $H_{(1)}$.

Remark 3. Note that we specify a selection matrix $\mathbb{C}_{K_0}$ in Theorem 3.3(i) (and Assumption A.5(i)) that is not needed if K_0 is fixed. When the dimension of $\beta^0_{(1)}$, namely, K_0, is diverging to infinity, we cannot derive the asymptotic normality of $\hat{\beta}_{(1)}$ directly. Instead, we follow the literature on inferences with a diverging number of parameters (see, e.g., Fan and Peng (2004), Lam and Fan (2008), Lu and Su (2015)) and prove the asymptotic normality for any arbitrary linear combinations of elements of $\hat{\beta}_{(1)}$. To understand the results in Theorem 3.3, we consider an *oracle* who knows the exact number of factors and exact regressors that should be included in the panel regression model. In this case, one can consider the estimation of both the slope coefficients and the factors and factor loadings via the Gaussian QMLE method of Bai (2009). This one-step oracle estimator is asymptotically efficient under Gaussian errors and some other conditions. Ideally, one can consider a one-step SCAD or Bridge-type PLS regression where the penalty terms on both β and λ (or F) are added to $\mathcal{L}^0_{NT}(\beta,\lambda,F)$ defined in (2.4) instead of $\mathcal{L}^0_{NT}(\beta,\lambda,\hat{F})$. We conjecture that such a one-step PLS estimator is as efficient as the one-step oracle estimator. Nevertheless, because we observe neither F nor λ and some identification restrictions on F and λ are required, it is very challenging to study the asymptotic properties of such a one-step PLS estimator.[2] For this reason, we compare our estimator with an alternative two-step estimator that is obtained by a second step augmented regression with estimated factors obtained using Bai's (2009) PCA-based QMLE method from a first-step estimation; see, e.g., Kapetanios and Pesaran (2007) and GHS. This two-step augmented estimator is only as efficient as the one-step QMLE under some restrictive assumptions and has more bias terms to be corrected otherwise. But after the bias correction, it is asymptotically equivalent to the bias-corrected one-step QMLE estimator. See also Remark 7 below.

Specifically, let $\bar{\beta}_{(1)}$ and $\bar{\lambda}_{i(1)}$ denote the least squares (LS) estimates of $\beta^0_{(1)}$ and $\lambda^*_{i(1)}=H^{+}_{(1)}\lambda^0_i$ in the following augmented panel regression

$$Y_{it}=\beta^{0\prime}_{(1)}X_{it(1)}+\lambda^{*\prime}_{i(1)}\hat{F}_{t(1)}+\epsilon_{it}, \tag{3.1}$$

where ϵ_{it} is the new error term that takes into account the estimation error from the first stage estimation. Then we can readily show that

$$\bar{\beta}_{(1)}=\hat{D}^{-1}_{\hat{F}_{(1)}}\frac{1}{NT}\sum_{i=1}^{N}X'_{i(1)}M_{\hat{F}_{(1)}}Y_{i(1)},\ \bar{\lambda}_{i(1)}=\hat{\Sigma}^{-1}_{\hat{F}_{(1)}}\frac{1}{T}\sum_{t=1}^{T}\hat{F}_{t(1)}(Y_{it}-X'_{it(1)}\bar{\beta}_{(1)}),$$

where $\hat{D}_{\hat{F}_{(1)}}\equiv(NT)^{-1}\sum_{i=1}^{N}X'_{i(1)}M_{\hat{F}_{(1)}}X_{i(1)}$ and $\hat{\Sigma}_{\hat{F}_{(1)}}\equiv\frac{1}{T}\sum_{t=1}^{T}\hat{F}_{t(1)}\hat{F}'_{t(1)}$. As demonstrated in the proof of Theorem 3.3, Assumption A.6 is essential to ensure that $(\hat{\beta}_{(1)},\hat{\lambda}_{i(1)})$ is asymptotically equivalent to $(\bar{\beta}_{(1)},\bar{\lambda}_{i(1)})$ in the sense that they share the same first order asymptotic distribution. For this reason,

[2] Bai and Liao (2013) propose a one-step shrinkage estimator for a pure factor model where the error terms are heteroskedastic and cross-sectionally correlated but exhibit a conditionally sparse covariance matrix. Under the assumption that the true number of factors is *known*, they establish the consistency of their estimator but state that deriving the limiting distribution is technically difficult.

we say that our estimator $(\hat{\beta}_{(1)}, \hat{\lambda}_{i(1)})$ is as oracle efficient as a two-step augmented estimator by knowing the exact number of factors and regressors.

Remark 4. Despite the oracle property of $\hat{\beta}_{(1)}$, it possesses four bias terms that have to be corrected in practice in order to ensure its $\sqrt{NT}$-consistency and zero-mean asymptotic normality. Interestingly, under a different set of assumptions, GHS establish formal asymptotics for the factor-augmented panel regressions in the case of fixed K_0. They show that the replacement of the unobservable factor F_t^0 by the PCA estimate $\hat{F}_{t(1)}$ in (3.1) does not affect the limiting distribution of the LS estimates of $\beta_{(1)}^0$ under four key conditions: (i) $T/N \to 0$ and $N/T^3 \to 0$, (ii) there is no dynamic lagged dependent variable in the regression, (iii) X_i also possesses a factor structure: $X_i = F^X \lambda_i^X + V_i$, and the estimated factors associated with X_i are also included into the augmented regression, and (iv) the exact number of factors and the exact regressors that should be included in the model are known.[3] Note that we relax all the four assumptions in this paper. We relax condition (iv) by considering the shrinkage estimation. Under condition (i), both $\mathbb{B}_{1NT}$ and $\mathbb{B}_{3NT}$ are $o_P(1)$. Under condition (iii), F^X is a submatrix of F^0 so that $F^{X\prime} M_{F^0} = 0$ and the factor component of $X_{i(1)}$ does not contribute to $\mathbb{B}_{2NT}$; under GHS's conditions on V_i, ε_i, F_t^0 and λ_i^0 (see their Assumptions A(v)-(viii)), the error component of $X_{i(1)}$ does not contribute to $\mathbb{B}_{2NT}$ either. That is, $\mathbb{B}_{2NT}$ is asymptotically negligible under their conditions. To understand the sources of asymptotic bias and variance of our estimator, we consider the following expansion used in the proof of Theorem 3.3:

$$\begin{aligned}\sqrt{NT}\mathbb{C}_{K_0}\left(\hat{\beta}_{(1)} - \beta_{(1)}^0\right) &= \mathbb{C}_{K_0}\hat{D}_{\hat{F}_{(1)}}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^N X_{i(1)}' M_{F^0}\varepsilon_i \\ &\quad + \mathbb{C}_{K_0}\hat{D}_{\hat{F}_{(1)}}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^N X_{i(1)}' M_{\hat{F}_{(1)}}(F_{(1)}^* - \hat{F}_{(1)})\lambda_{i(1)}^* \\ &\quad + \mathbb{C}_{K_0}\hat{D}_{\hat{F}_{(1)}}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^N X_{i(1)}'\left(M_{\hat{F}_{(1)}} - M_{F_{(1)}^*}\right)\varepsilon_i + o_P(1) \\ &\equiv S_{1NT} + S_{2NT} + S_{3NT} + o_P(1), \text{ say.}\end{aligned} \tag{3.2}$$

S_{1NT} is present even if one observes F_t^0 (in which case $\hat{D}_{\hat{F}_{(1)}}$ is replaced by $\hat{D}_{F^0}$). We show that S_{2NT} contributes to both the asymptotic bias and variance whereas S_{3NT} only contributes to the asymptotic bias:

$$S_{2NT} = \mathbb{C}_{K_0}\hat{D}_{\hat{F}_{(1)}}^{-1}\left(\mathbb{B}_{1NT} - \mathbb{B}_{2NT} - \mathbb{V}_{1NT} + \mathbb{V}_{2NT}\right) + o_P(1), \text{ and } S_{3NT} = -\mathbb{C}_{K_0}\hat{D}_{\hat{F}_{(1)}}^{-1}\mathbb{B}_{3NT} + o_P(1),$$

where $\mathbb{V}_{1NT} \equiv (NT)^{-1/2}\sum_{i=1}^N \mathcal{X}_{i1NT}' M_{F^0}\varepsilon_i$, $\mathbb{V}_{2NT} \equiv C_{NT}'\sqrt{NT}(\tilde{\beta}^c - \beta^0)$, and $\mathcal{X}_{i1NT}' = \frac{1}{N}\sum_{j=1}^N \lambda_i^{0\prime}[T^{-1}F^{0\prime}\tilde{F}_{(1)}]\lambda_{j(1)}^* X_{j(1)}'$. In general, the parameter estimation error plays an important role. Nevertheless, under GHS's key condition (ii) in conjunction with some other regularity conditions specified in their Assumption A, one can show that $\|\mathbb{V}_{1NT}\| = o_P(1)$ and $\|C_{NT}\|_{\text{sp}} = o_P(1)$ (and hence $C_0 = 0$ in our Assumption

[3] Condition (i) is explicitly mentioned in GHS. Lagged dependent variables are ruled out by the second part of Assumption B in their paper. The first part of (iii) is explicitly assumed in their equation (3) and the second part is implicitly assumed because the factors in their equation (6) include the maximal common factor set of the observable variables (Y_i, X_i). (iv) is also implicitly assumed in their paper.

A.4(ii) and $\|\mathbb{V}_{2NT}\| = o_P(1)$ under our Assumption A.1(i)). In this case, only S_{1NT} contributes to the asymptotic variance of their augmented estimator of $\beta^0_{(1)}$ and both S_{2NT} and S_{3NT} are asymptotically negligible under GHS's key conditions (i), (iii) and (iv). If their key condition (ii) is also satisfied, one can show that S_{1NT} converges to a zero-mean normal distribution; otherwise, one has to consider bias-correction as in Moon and Weidner (2014a).

Remark 5. The presence of $\mathbb{B}_{4NT}$ and the complicated structure of $\bar{Z}_i$ in Assumption A.5(i) are mainly due to the allowance of lagged dependent variables because X_{it} can be correlated with ε_{is} for $t > s$.[4] In this case, S_{1NT} is not centered around zero asymptotically, whereas both $\mathbb{V}_{1NT}$ and $\mathbb{V}_{2NT}$ are centered around 0 asymptotically.[5] We have to decompose S_{1NT} into an asymptotic bias term (which is associated with $\mathbb{B}_{4NT}$) and an asymptotic variance term (which enters $\bar{Z}_i$ via $X_{i(1)} - P_{F^0}E_{\mathcal{D}}(X_{i(1)})$):

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} X'_{i(1)}M_{F^0}\varepsilon_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[X'_{i(1)}\varepsilon_i - E_{\mathcal{D}}(X'_{i(1)})P_{F^0}\varepsilon_i\right] - \mathbb{B}_{4NT}.$$

We can find primitive conditions to ensure that the first term in the last expression converges to a zero mean normal distribution, the conditional expectation $\bar{\mathbb{B}}_{4NT}$ of the second term given $\mathcal{D}$ contributes to the asymptotic bias which can be corrected, and $\mathbb{B}_{4NT} - \bar{\mathbb{B}}_{4NT}$ is asymptotically negligible. For further details, see the proofs of Theorem 3.3 and Corollary 3.4 in the appendix and the supplementary appendix, respectively.

Remark 6. Now we consider some special cases where the formulae for the asymptotic bias and variance terms can be simplified.

1. If all regressors are strictly exogenous as in Pesaran (2006), Bai (2009), and GHS, then one can set $\mathbb{B}_{4NT} = 0$ and $\bar{Z}_i \equiv M_{F^0}[X_{i(1)} - \mathcal{X}_{i1NT} + (X_i - \mathcal{X}_{i2NT})W_0^{-1}C_0]$ in Assumption A.5(i), where $\mathcal{X}_{i1NT} = \frac{1}{N}\sum_{j=1}^{N}\lambda_i^{0\prime}[T^{-1}F^{0\prime}\tilde{F}_{(1)}]\lambda^*_{j(1)}X_{j(1)}$ and $\mathcal{X}_{i2NT} = \frac{1}{N}\sum_{j=1}^{N}\lambda_i^{0\prime}\left(N^{-1}\lambda^{0\prime}\lambda^0\right)^{-1}\lambda_j^0 X_j$. In short, there is no need to consider conditioning on the "exogenous" set of factors and factor loadings.

2. If in addition, X_i also follows a factor structure as in GHS, then there is no need to correct $\mathbb{B}_{2NT}$ and $\mathbb{B}_{4NT}$ under the conditions specified in GHS, and one can reset $\bar{Z}_i \equiv M_{F^0}X_{i(1)}$ in Assumption A.5(i).

3. If in addition, $T/N \to 0$, there is no need to correct any bias term.

Note that we present Theorem 3.3 under a set of fairly general and high level assumptions. To estimate the asymptotic bias and variance, one generally needs to add more specific assumptions as in Bai (2009). In the supplementary appendix, we specify a set of assumptions (Assumptions B.1-B.2) that ensure all the high level conditions specified in Assumptions A.1(vi)-(vii), A.2(ii)-(iii), A.4(iii)-(iv) and A.5(i)-(ii) to be satisfied. Note that Assumption B.1(i) relies on the key notion of *conditional strong mixing* that

[4] In the absense of lagged dependent variables, one can simply combine S_{1NT} with $D^{-1}_{\tilde{F}_{(1)}}(-\mathbb{V}_{1NT} + \mathbb{V}_{2NT})$ to obtain the asymptotic distribution without GHS's key conditions (i), (iii) and (iv).

[5] $\mathbb{V}_{1NT}$ is centered around 0 asymptotically because $\mathcal{X}_{i1NT}$ is defined as a weighted average of $X_{j(1)}$ which asymptotically smooths out the endogenous component of X_{jt}; $\mathbb{V}_{2NT}$ is also asymptotically centered around 0 because of the adoption of a bias-corrected estimate $\tilde{\beta}^c$ in its definition.

has recently been introduced by Prakasa Rao (2009) and Roussas (2008) and applied to the econometrics literature by Su and Chen (2013) and Moon and Weidner (2014b). Assumptions B.1-B.2 are also used to establish the consistency of the asymptotic bias and variance estimates.

In particular, under the martingale difference sequence (m.d.s.) condition in Assumption B.2, we have

$$\Theta_{NT} = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{it}^2 \bar{Z}_{it} \bar{Z}_{it}', \ \Theta_{i,F_{(1)}^*} = \lim_{(N,T)\to\infty} \Theta_{iNT,F_{(1)}^*}, \ \Theta_{iNT,F_{(1)}^*} \equiv \frac{1}{T} \sum_{t=1}^{T} E\left[F_{t(1)}^* F_{t(1)}^{*\prime} \varepsilon_{it}^2\right],$$

where $\bar{Z}_{it}'$ denotes the tth row of $\bar{Z}_i$. One can consistently estimate $\Theta_{i,F_{(1)}^*}$ by $\hat{\Theta}_{i,F_{(1)}^*} = \frac{1}{T}\sum_{t=1}^{T} \hat{F}_{t(1)} \hat{F}_{t(1)}' \hat{\varepsilon}_{it}^2$, where $\hat{\varepsilon}_{it} = Y_{it} - X_{it(1)}' \hat{\beta}_{(1)} - \hat{\lambda}_{i(1)}' \hat{F}_{t(1)}$. Below we focus on inferential theory for $\beta_{(1)}^0$.

Let $\hat{\Psi}_{NT} \equiv$diag$(\hat{\psi}_{1T}, ..., \hat{\psi}_{NT})$ and $\hat{\Phi}_{NT} \equiv$diag$(\hat{\varphi}_{1N}, ..., \hat{\varphi}_{TN})$ where $\hat{\psi}_{iT} \equiv T^{-1}\sum_{t=1}^{T} \hat{\varepsilon}_{it}^2$ and $\hat{\varphi}_{tN} \equiv N^{-1}\sum_{i=1}^{N} \hat{\varepsilon}_{it}^2$. Let $\hat{Z}_i \equiv X_{i(1)} - P_{\hat{F}_{(1)}} X_{i(1)} - M_{\hat{F}_{(1)}} \hat{\mathcal{X}}_{i1NT} + [X_i - P_{\hat{F}_{(1)}} X_i - M_{\hat{F}_{(1)}} \hat{\mathcal{X}}_{i2NT}] \hat{W}_{NT}^{-1} \hat{C}_{NT}$, where $\hat{\mathcal{X}}_{i1NT} \equiv \frac{1}{N}\sum_{j=1}^{N} \hat{\lambda}_{i(1)}' [T^{-1} \hat{F}_{(1)}' \tilde{F}_{(1)}] \hat{\lambda}_{j(1)} X_{j(1)}$, and $\hat{\mathcal{X}}_{i2NT} \equiv \frac{1}{N}\sum_{j=1}^{N} \hat{\lambda}_{i(1)}' [N^{-1} \hat{\lambda}_{(1)}' \hat{\lambda}_{(1)}]^{-1} \hat{\lambda}_{j(1)} X_j$, $\hat{C}_{NT}' \equiv (NT)^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N} \hat{\lambda}_{j(1)}' \hat{F}_{(1)}' \tilde{F}_{(1)} \hat{\lambda}_{i(1)} X_{i(1)}' M_{\hat{F}_{(1)}} X_j$, $\hat{W}_{NT} \equiv \frac{1}{NT}\sum_{i=1}^{N} \hat{\bar{X}}_i' M_{\hat{F}_{(1)}} \hat{\bar{X}}_i$, and $\hat{\bar{X}}_i \equiv X_i - \hat{\mathcal{X}}_{i2NT}$. Note that we can write the kth elements of $\mathbb{B}_{3NT}$ and $\bar{\mathbb{B}}_{4NT} \equiv E_{\mathcal{D}}(\mathbb{B}_{4NT})$ respectively as

$$\mathbb{B}_{3NT,k} \equiv (NT)^{-3/2} \operatorname{tr}\left[(F_{(1)}^{*\prime} F_{(1)}^*)^{-1} \tilde{F}_{(1)}' F^0 \lambda^{0\prime} \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mathbf{X}_k F_{(1)}^*\right] \text{ and } \bar{\mathbb{B}}_{4NT,k} \equiv (NT)^{-1/2} \operatorname{tr}\left[P_{F^0} E_{\mathcal{D}}(\boldsymbol{\varepsilon}' \mathbf{X}_k)\right].$$

We propose to estimate the bias and variance terms as follows:

$$\begin{aligned}
\hat{\mathbb{B}}_{1NT} &\equiv N^{-5/2} T^{-3/2} \sum_{i=1}^{N} X_{i(1)}' \hat{F}_{(1)} (\hat{F}_{(1)}' \hat{F}_{(1)})^{-1} \tilde{F}_{(1)}' \hat{F}_{(1)} \hat{\lambda}_{(1)}' \hat{\Psi}_{NT} \hat{\lambda}_{(1)} \hat{F}_{(1)}' \tilde{F}_{(1)} \hat{\lambda}_{i(1)}, \\
\hat{\mathbb{B}}_{2NT} &\equiv N^{-1/2} T^{-3/2} \sum_{i=1}^{N} X_{i(1)}' M_{\hat{F}_{(1)}} \hat{\Phi}_{NT} \tilde{F}_{(1)} \hat{\lambda}_{i(1)}, \\
\hat{\mathbb{B}}_{3NT,k} &\equiv N^{-3/2} T^{-1/2} \operatorname{tr}\left[(\hat{F}_{(1)}' \hat{F}_{(1)})^{-1} \tilde{F}_{(1)}' \hat{F}_{(1)} \hat{\lambda}_{(1)}' \hat{\Psi}_{NT} \mathbf{X}_k \hat{F}_{(1)}\right] \text{ for } k = 1, ..., K_0, \\
\hat{\mathbb{B}}_{4NT,k} &\equiv \frac{1}{\sqrt{NT}} \operatorname{tr}\left[P_{\hat{F}_{(1)}} \left(\hat{\boldsymbol{\varepsilon}}' \mathbf{X}_k\right)^{\text{trunc}}\right] \text{ for } k = 1, ..., K_0, \\
\hat{\Theta}_{NT} &\equiv \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\varepsilon}_{it}^2 \hat{Z}_{it} \hat{Z}_{it}',
\end{aligned}$$

where $A^{\text{trunc}} \equiv \sum_{t=1}^{T-M} \sum_{s=t+1}^{t+M} A_{ts}^*$ for any $T \times T$ matrix $A = (A_{ts})$ and A_{ts}^* is a $T \times T$ matrix with (t,s)th element given by A_{ts} and zeros elsewhere, and $\hat{Z}_{it}'$ denotes the tth row of $\hat{Z}_i$. Let $\hat{\mathbb{B}}_{3NT} \equiv (\hat{\mathbb{B}}_{3NT,1}, ..., \hat{\mathbb{B}}_{3NT,K_0})'$ and $\hat{\mathbb{B}}_{4NT} \equiv (\hat{\mathbb{B}}_{4NT,1}, ..., \hat{\mathbb{B}}_{4NT,K_0})'$. We define the bias-corrected adaptive group Lasso estimator of $\beta_{(1)}^0$ as

$$\hat{\beta}_{(1)}^c = \hat{\beta}_{(1)} - (NT)^{-1/2} \hat{D}_{\hat{F}_{(1)}}^{-1} (\hat{\mathbb{B}}_{1NT} - \hat{\mathbb{B}}_{2NT} - \hat{\mathbb{B}}_{3NT} - \hat{\mathbb{B}}_{4NT}).$$

The following corollary establishes the asymptotic distribution of $\hat{\beta}_{(1)}^c$.

Corollary 3.4 *Suppose Assumptions A.1(i), (v), (viii), A.3, A.4(i)-(ii), A.6, and B.1-B.2 hold. Let* $\hat{\mathbb{V}}_{NT} = \hat{D}_{\hat{F}_{(1)}}^{-1} \hat{\Theta}_{NT} \hat{D}_{\hat{F}_{(1)}}^{-1}$. *Then* $\mathbb{C}_{K_0} \sqrt{NT} (\hat{\beta}_{(1)}^c - \beta_{(1)}^0) \xrightarrow{D} N\left(0, \lim_{(N,T)\to\infty} \mathbb{C}_{K_0} \mathbb{V}_{NT} \mathbb{C}_{K_0}'\right)$ *and* $\mathbb{C}_{K_0}(\hat{\mathbb{V}}_{NT} - \mathbb{V}_{NT}) \mathbb{C}_{K_0}' = o_P(1)$

Remark 7. The proof of the above corollary is quite involved and we delegate it to the supplementary appendix. If only strictly exogenous regressors are present in the model, following Remark 6, we can set $\hat{\mathbb{B}}_{4NT} = 0$ and redefine $\hat{Z}_i \equiv M_{\hat{F}_{(1)}}[X_{i(1)} - \hat{\mathcal{X}}_{i1NT} + (X_i - \hat{\mathcal{X}}_{i2NT})\hat{W}_{NT}^{-1}\hat{C}_{NT}]$ to be used in the variance estimation. When other conditions are also satisfied, both the bias and variance estimates can be further simplified with obvious modifications according to Remark 6. It is worth mentioning that our bias-corrected estimator is asymptotically equivalent to Moon and Weidner's (2014b) bias-corrected estimator in the case where all regressors are *relevant* (i.e., there is no selection of regressors) and K_0 is fixed. In the presence of *irrelevant* regressors, the variance-covariance matrix for our shrinkage estimator of the non-zero coefficients is smaller than that of Moon and Weidner's estimator.

Remark 8. Belloni and Chernozhukov (2013) study post-model selection estimators which apply ordinary least squares to the model selected by first-step penalized estimators and show that the post Lasso estimators perform at least as well as Lasso in terms of the rate of convergence and have the advantage of having a smaller bias. After we apply our adaptive group Lasso procedure, we can re-estimate the panel data model based on the selected regressors and number of factors and the QMLE method of Bai (2009) or Moon and Weidner (2014a, 2014b). We will compare the performance of these post-Lasso estimators with the Lasso estimators through simulations.

Remark 9. Note that our asymptotic results are "pointwise" in the sense that the unknown parameters are treated as fixed. The implication is that in finite samples, the distributions of our estimators can be quite different from the normal, as discussed in Leeb and Pöscher (2005, 2008, 2009) and Schneider and Pöscher (2009). This is a well-known challenge in the literature of model selection no matter whether the selection is based on a information criterion or Lasso-type technique. Despite its importance, developing a thorough theory on uniform inference is beyond the scope of this paper.

Remark 10. As a referee kindly points out, our procedure does not take into account the possible correlation in ε_{it} and it may not work well in the case of strong serial correlation like Bai and Ng's (2002) information criterion. Suppose that the error term has an AR(1) structure: $\varepsilon_{it} = \rho^0 \varepsilon_{it-1} + e_{it}$, where $\{e_{it}, t \geq 1\}$ is a white noise for each i. Then one can transform the original model in (2.1) via the Cochrane and Orcutt's (1949) procedure to obtain

$$Y_{it} - \rho^0 Y_{it-1} = \beta^{0\prime} \left(X_{it} - \rho^0 X_{it-1} \right) + \lambda_i^{0\prime} \breve{F}_t^0 + e_{it}, \tag{3.3}$$

where $\breve{F}_t^0 = \left(F_t^0 - \rho^0 F_{t-1}^0 \right)$. We propose the following two-stage estimator:

Stage 1: Obtain the residuals $\hat{\varepsilon}_{it}$ using the largest model (i.e., $r = R$ and including all regressors) and let $\hat{\rho}$ be the OLS estimator of ρ by regressing $\hat{\varepsilon}_{it}$ on $\hat{\varepsilon}_{it-1}$.

Stage 2: Apply our Lasso method to the following transformed model:

$$(Y_{it} - \hat{\rho} Y_{it-1}) = \beta^{0\prime} (X_{it} - \hat{\rho} X_{it-1}) + \lambda_i^{0\prime} \breve{F}_t^0 + \eta_{it}, \tag{3.4}$$

where η_{it} is a new error term that incorporates both the original error term e_{it} and the estimation error due to the replacement of ρ^0 by $\hat{\rho}$. Simulations demonstrate such a method works fairly well in the case of serially correlated errors.

3.4 Choosing the Tuning Parameter γ

Let $\mathcal{S}_{\beta}(\gamma)$ and $\mathcal{S}_{\lambda}(\gamma)$ denote the index set of nonzero elements in $\hat{\beta}(\gamma)$ and nonzero columns in $\hat{\lambda}(\gamma)$, respectively. Let $\mathcal{S}(\gamma)=\mathcal{S}_{\beta}(\gamma)\times\mathcal{S}_{\lambda}(\gamma)$. Let $|\mathcal{S}|$ denote the cardinality of the index set $\mathcal{S}$. We propose to select the tuning parameter $\gamma=(\gamma_1,\gamma_2)$ by minimizing the following information criterion:

$$IC(\gamma)=\hat{\sigma}^2(\gamma)+\rho_{1NT}\left|\mathcal{S}_{\beta}(\gamma)\right|+\rho_{2NT}N\left|\mathcal{S}_{\lambda}(\gamma)\right|, \tag{3.5}$$

where $\hat{\sigma}^2(\gamma)=\mathcal{L}_{NT}(\hat{\beta}(\gamma),\hat{\lambda}(\gamma))$. Similar information criteria are proposed by Wang et al. (2007) and Liao (2013) for shrinkage estimation in different contexts.

Let $\mathcal{S}_{F,\beta}=\{1,...,K\}$ and $\mathcal{S}_{T,\beta}=\{1,...,K_0\}$ denote the index sets for the full set of covariates and the (true) set of relevant covariates in X_{it}, respectively. Similarly, $\mathcal{S}_{F,\lambda}=\{1,...,R\}$ and $\mathcal{S}_{T,\lambda}=\{1,...,R_0\}$ denote the index sets for the full set of factors and the (true) set of relevant factors in F_t, respectively. Let $\mathcal{S}=(\mathcal{S}_{\beta},\mathcal{S}_{\lambda})$ be an arbitrary index set with $\mathcal{S}_{\beta}=\{j_1,...,j_{K^*}\}\subset\mathcal{S}_{F,\beta}$ and $\mathcal{S}_{\lambda}=\{j_1,...,j_{R^*}\}\subset\mathcal{S}_{F,\lambda}$ where $0\leq K^*\leq K$ and $0\leq R^*\leq R$. Consider a candidate model with regressor index $\mathcal{S}_{\beta}$ and factor index $\mathcal{S}_{\lambda}$. Then any candidate model with either $\mathcal{S}_{\beta}\not\supseteq\mathcal{S}_{T,\beta}$ or $\mathcal{S}_{\lambda}\not\supseteq\mathcal{S}_{T,\lambda}$ is referred to as an *under-fitted* model in the sense that it misses at least one important covariate or factor. Similarly, any candidate model with $\mathcal{S}_{\beta}\supset\mathcal{S}_{T,\beta}$, $\mathcal{S}_{\lambda}\supset\mathcal{S}_{T,\lambda}$ and either $\mathcal{S}_{\beta}\neq\mathcal{S}_{T,\beta}$ or $\mathcal{S}_{\lambda}\neq\mathcal{S}_{T,\lambda}$ (i.e., $|\mathcal{S}_{\beta}|+|\mathcal{S}_{\lambda}|>|\mathcal{S}_{T,\beta}|+|\mathcal{S}_{T,\lambda}|$) is referred as an *over-fitted* model in the sense that it contains not only all relevant covariates and factors but also at least one irrelevant covariate or factor.

Denote $\Omega_1=[0,\gamma_{1\max}]$ and $\Omega_2=[0,\gamma_{2\max}]$, two bounded intervals in $\mathbb{R}^+$, where the potential tuning parameters γ_{1NT} and γ_{2NT} take values, respectively. Here we suppress the dependence of Ω_1, Ω_2, $\gamma_{1\max}$ and $\gamma_{2\max}$ on (N,T). We divide $\Omega=\Omega_1\times\Omega_2$ into three subsets Ω_0, Ω_- and Ω_+ as follows

$$\begin{aligned}
\Omega_0 &= \left\{\gamma\in\Omega:\mathcal{S}_{\beta}(\gamma)=\mathcal{S}_{T,\beta}\text{ and }\mathcal{S}_{\lambda}(\gamma)=\mathcal{S}_{T,\lambda}\right\},\\
\Omega_- &= \left\{\gamma\in\Omega:\mathcal{S}_{\beta}(\gamma)\not\supseteq\mathcal{S}_{T,\beta}\text{ or }\mathcal{S}_{\lambda}(\gamma)\not\supseteq\mathcal{S}_{T,\lambda}\right\},\\
\Omega_+ &= \left\{\gamma\in\Omega:\mathcal{S}_{\beta}(\gamma)\supset\mathcal{S}_{T,\beta},\,\mathcal{S}_{\lambda}(\gamma)\supset\mathcal{S}_{T,\lambda},\text{ and }|\mathcal{S}_{\beta}|+|\mathcal{S}_{\lambda}|>|\mathcal{S}_{T,\beta}|+|\mathcal{S}_{T,\lambda}|\right\}.
\end{aligned}$$

Clearly, Ω_0, Ω_- and Ω_+ denote the three subsets of Ω in which the true, under- and over-fitted models can be produced.

For any $\mathcal{S}=\mathcal{S}_{\beta}\times\mathcal{S}_{\lambda}$ with $\mathcal{S}_{\beta}=\{j_1,...,j_{|\mathcal{S}_{\beta}|}\}\subset\mathcal{S}_{F,\beta}$ and $\mathcal{S}_{\lambda}=\{l_1,...,l_{|\mathcal{S}_{\lambda}|}\}\subset\mathcal{S}_{F,\lambda}$, we use $\beta_{\mathcal{S}_{\beta}}=(\beta_{j_1},...,\beta_{j_{|\mathcal{S}_{\beta}|}})'$ to denote an $|\mathcal{S}_{\beta}|\times 1$ subvector of β and $\lambda_{\mathcal{S}_{\lambda}}=(\lambda_{l_1},...,\lambda_{l_{|\mathcal{S}_{\lambda}|}})$ to denote an $N\times|\mathcal{S}_{\lambda}|$ submatrix of λ. Similarly, $X_{it,\mathcal{S}_{\beta}}$ and $\hat{F}_{t,\mathcal{S}_{\lambda}}$ denote the $|\mathcal{S}_{\beta}|\times 1$ subvector of X_{it} and $|\mathcal{S}_{\lambda}|\times 1$ subvector of $\hat{F}_t$ according to $\mathcal{S}$. Let $\hat{\beta}_{\mathcal{S}_{\beta}}$ and $\hat{\lambda}_{\mathcal{S}_{\gamma}}$ denote the ordinary least squares (OLS) estimators of $\beta_{\mathcal{S}_{\beta}}$ and $\lambda_{\mathcal{S}_{\lambda}}$, respectively, by regressing Y_{it} on $X_{it,\mathcal{S}_{\beta}}$ and $\hat{F}_{t,\mathcal{S}_{\lambda}}$. Define

$$\hat{\sigma}^2_{\mathcal{S}}=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Y_{it}-\hat{\beta}'_{\mathcal{S}_{\beta}}X_{it,\mathcal{S}_{\beta}}-\hat{\lambda}'_{i,\mathcal{S}_{\lambda}}\hat{F}_{t,\mathcal{S}_{\lambda}}\right)^2, \tag{3.6}$$

where $\hat{\lambda}'_{i,\mathcal{S}_{\lambda}}$ denotes the ith row of $\hat{\lambda}_{\mathcal{S}_{\lambda}}$. Let $\mathcal{S}_T=\mathcal{S}_{T,\beta}\times\mathcal{S}_{T,\lambda}$. One can readily show that $\hat{\sigma}^2_{\mathcal{S}_T}\xrightarrow{P}\sigma^2_{\mathcal{S}_T}\equiv\lim_{(N,T)\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left(\varepsilon_{it}^2\right)$ under Assumptions A.1-A.2.

To proceed, we add the following two assumptions.

Assumption A.7 For any $\mathcal{S} = \mathcal{S}_\beta \times \mathcal{S}_\lambda$ with either $\mathcal{S}_\beta \not\supseteq \mathcal{S}_{T,\beta}$ or $\mathcal{S}_\lambda \not\supseteq \mathcal{S}_{T,\lambda}$, there exists $\sigma^2_{\mathcal{S}}$ such that $\hat{\sigma}^2_{\mathcal{S}} \xrightarrow{P} \sigma^2_{\mathcal{S}} > \sigma^2_{\mathcal{S}_T}$.

Assumption A.8 As $(N, T) \to \infty$, $\rho_{1NT} K_0 \to 0$, $\rho_{2NT} N \to 0$, $\rho_{1NT} \delta^2_{NT} \to \infty$, and $\rho_{2NT} N \delta^2_{NT} \to \infty$.

Assumption A.7 is intuitively clear. It requires that all under-fitted models yield asymptotic mean square errors that are larger than $\sigma^2_{\mathcal{S}_T}$, which is delivered by the true model. A.8 reflects the usual conditions for the consistency of model selection. The penalty coefficients ρ_{1NT} and ρ_{2NT} cannot shrink to zero either too fast or too slowly.

Let $\gamma^0_{NT} = \left(\gamma^0_{1NT}, \gamma^0_{2NT}\right)'$ where γ^0_{1NT} and γ^0_{2NT} satisfy the conditions on γ_{1NT} and γ_{2NT}, respectively in Assumptions A.3(ii)-(iii).

Theorem 3.5 *Suppose that Assumptions A.1, A.2, A.3(i), A.6(i), A.7 and A.8 hold. Then*

$$P\left(\inf_{\gamma \in \Omega_- \cup \Omega_+} IC(\gamma) > IC\left(\gamma^0_{NT}\right)\right) \to 1 \text{ as } (N, T) \to \infty.$$

Remark 11. Note that we do not impose Assumptions A.3(ii)-(iii), A.4, A.5 and A.6(ii) in the above theorem. Theorem 3.5 implies that the γ's that yield the over- or under-selected sets of regressors or number of factors fail to minimize the information criterion w.p.a.1. Consequently, the minimizer of $IC(\gamma)$ can only be the one that meets Assumptions A.3(ii)-(iii) so that both estimation and selection consistency can be achieved.

4 Monte Carlo Simulations

In this section, we conduct Monte Carlo simulations to evaluate the finite sample performance of our proposed adaptive group Lasso (agLasso) method.

4.1 Data Generating Processes

We consider the following data generating processes (DGPs):
DGP 1: $Y_{it} = \lambda^0_{i,1} F^0_{t,1} + \lambda^0_{i,2} F^0_{t,2} + \theta\varepsilon_{it}$,
DGP 2: $Y_{it} = \beta^0_1 X_{it,1} + \beta^0_2 X_{it,2} + \lambda^0_{i,1} F^0_{t,1} + \lambda^0_{i,2} F^0_{t,2} + \theta\varepsilon_{it}$, where $\left(\beta^0_1, \beta^0_2\right) = (1, 1)$,
DGP 3: $Y_{it} = \beta^0_1 X_{it,1} + \beta^0_2 Y_{it-1} + \lambda^0_{i,1} F^0_{t,1} + \lambda^0_{i,2} F^0_{t,2} + \theta\varepsilon_{it}$, where $\left(\beta^0_1, \beta^0_2\right) = (1, 0.25)$,
DGP 4: $Y_{it} = \beta^0_1 X_{it,1} + \beta^0_2 X_{it,2} + \beta^0_3 X_{it,3} + \beta^0_4 X_{it,4} + \beta^0_5 X_{it,5} + \lambda^0_{i,1} F^0_{t,1} + \lambda^0_{i,2} F^0_{t,2} + \theta\varepsilon_{it}$, where $\left(\beta^0_1, \beta^0_2, \beta^0_3, \beta^0_4, \beta^0_5\right) = (1, 1, 0, 0, 0)$,
DGP 5: $Y_{it} = \beta^0_1 X_{it,1} + \beta^0_2 Y_{it-1} + \beta^0_3 X_{it,2} + \beta^0_4 X_{it,3} + \beta^0_5 Y_{it-2} + \lambda^0_{i,1} F^0_{t,1} + \lambda^0_{i,2} F^0_{t,2} + \theta\varepsilon_{it}$, where $\left(\beta^0_1, \beta^0_2, \beta^0_3, \beta^0_4, \beta^0_5\right) = (1, 0.25, 0, 0, 0)$,
DGP 6: $Y_{it} = \sum_{k=1}^{K} \beta^0_k X_{it,k,} + \lambda^0_{i,1} F^0_{t,1} + \lambda^0_{i,2} F^0_{t,2} + \theta\varepsilon_{it}$, where $\left(\beta^0_1, \beta^0_2\right) = (1, 1)$, $\beta_k = 0$ for $k = 3, ..., K$, and $K = \lfloor 5(NT)^{1/5} \rfloor$.

In all the six DGPs, $\lambda^0_{i,1}$, $\lambda^0_{i,2}$, and ε_{it} are independent $N(0, 1)$ random variables. In DGPs 1, 2, 4 and 6, $F^0_{t,1}$ and $F^0_{t,2}$ are independently and standard normally distributed. In DGPs 3 and 5, we consider an AR(1) structure for the factors: $F^0_{t,1} = 0.5 F^0_{t-1,1} + \zeta_{t,1}$ and $F^0_{t,2} = 0.5 F^0_{t-1,2} + \zeta_{t,2}$, where $(\zeta_{t,1}, \zeta_{t,2})$ are independent $N(0, 1)$ random variables. $X_{it,k} = 0.25(\lambda^0_{i1} F^0_{t1} + \lambda^0_{i2} F^0_{t2}) + \xi_{it,k}$, where $\xi_{it,k}$

are IID $N(0,1)$ across both i and t for $k = 1, ..., 10$ and the rest $X_{it,k}$'s are independent $N(0,1)$ for $k = 11, ..., K$ (in DGP 6). We use θ to control for the signal-to-noise (SN) ratio, which is defined as $\text{Var}\left(\beta^{0\prime} X_{it} + \lambda_i^{0\prime} F_t^0\right) / \text{Var}(\theta\varepsilon_{it})$. For each DGP, we choose θ's such that the SN ratio equals 1.[6]

DGP 1 is a pure factor structure without any regressor X_{it}. DGPs 2 and 3 are static and dynamic panel structures with interactive fixed effects, respectively. DGP 4 is identical to DGP 2 except that in DGP 4 we include three more irrelevant regressors: $X_{it,3}$, $X_{it,4}$ and $X_{it,5}$. Hence, in DGP 4, we consider both the selection of the regressors and determination of the number of factors, while in DGP 2 we only consider the latter. DGP 5 is identical to DGP 3, except that DGP 5 includes three irrelevant regressors: $X_{it,2}$, $X_{it,3}$ and Y_{it-2}. Thus, we select both the regressors and number of factors in DGP 5. DGP 6 is identical to DGPs 2 and 4 except that we consider a model with a growing number of regressors (K), where $K = \lfloor 5(NT)^{1/5} \rfloor$ and $\lfloor \cdot \rfloor$ denotes the integer part of $\cdot$. Note that in this model, K can be quite large, e.g., $K = 25$ when $N = 60$ and $T = 60$.

The true number of factors is 2 in all the above six DGPs. In our simulations, we assume that we do not know the true number of factors. We consider different combinations of (N, T) : $(20, 20)$, $(40, 40)$, $(20, 60)$, $(60, 20)$ and $(60, 60)$. The number of replications is 250.

4.2 Implementation

One of the important steps in our method is to choose the tuning parameters γ_{1NT} and γ_{2NT}. Following our theoretical arguments above, we use the information criterion in (3.5). Let s_Y^2 denote the sample variance of Y_{it}. For DGPs 1-3 where we only choose the number of factors, we set $\rho_{1NT} = 0$ and $\rho_{2NT} = s_Y^2 \ln(\delta_{NT}) / (N \min(N, T))$. For DGPs 4-6, we set $\rho_{1NT} = 0.05 s_Y^2 \ln(\delta_{NT}) / \min(N, T)$ and $\rho_{2NT} = s_Y^2 \ln(\delta_{NT}) / (N \min(N, T))$.[7] In DGPs 1-3, we only select factors, hence we let $\gamma_{1NT} = 0$ and choose γ_{2NT} from the set: $\left\{C_\gamma s_Y^2 \tau_1^{\kappa_2} (NT)^{-1/2} (\ln(NT))^{-1}\right\}$, where C_γ are 50 constants that increase geometrically from 0.01 to 25, i.e., $C_\gamma = 0.01$, 0.014, ..., 18.045, and 25. For DGPs 4-5, we let the candidate set of $(\gamma_{1NT}, \gamma_{2NT})$ be the Cartesian product: $\{C_\gamma s_Y^2 (NT)^{-1/2} (\ln(NT))^{-1}\} \times \{C_\gamma s_Y^2 \tau_1^{\kappa_2} (NT)^{-1/2} (\ln(NT))^{-1}\}$, where C_γ are 25 constants that increase geometrically from 0.01 to 25.[8] We set $\kappa_1 = \kappa_2 = 2$ in all cases.

We also consider choices of $(\gamma_{1NT}, \gamma_{2NT})$ based on a "rule of thumb" for DGPs 2-6:

$$(\gamma_{1NT}, \gamma_{2NT}) = c \cdot \left(s_Y^2 (NT)^{-1/2} (\ln(NT))^{-1}, \quad s_Y^2 \tau_1^{\kappa_2} (NT)^{-1/2} (\ln(NT))^{-1} \right)$$

where c is a constant and we use the fact that N and T pass to infinity at the same rate and $K_0 = 2$ is fixed in DGPs 2-6. Of course, we reset $\gamma_{1NT} = 0$ for DGP 1. We consider three values for c : 0.5, 1 and 2.

[6] The results for SN being 2 are reported in an early version of this paper and available upon request.

[7] A natural BIC-type choices of ρ_{1NT} and ρ_{1NT} that satisfy Assumption A.8 would be $\rho_{1NT} = \frac{s_Y^2 \ln(\delta_{NT})}{\delta_{NT}^2}$ and $\rho_{2NT} = \frac{s_Y^2 \ln(\delta_{NT})}{N \delta_{NT}^2}$. Neverthess, in practice $\ln(\delta_{NT}) \delta_{NT}^{-2} = \ln\left(\min(N^{1/2}, T^{1/2})\right) \max\left(N^{-1}, T^{-1}\right)$ is quite big in magnitude in comparison with the usual BIC tuning coefficient $\ln(NT)/(NT)$ as NT denotes the total number of observations in our panel data model. We find that through simulations that a downward adjustment of the above ρ_{1NT} by a scale of 1/10 would enhance the finite sample performance of the proposed IC. That is why we choose to use $\rho_{1NT} = 0.05 \frac{s_Y^2 \ln(\delta_{NT})}{\delta_{NT}^2}$ in our simulations and applications.

[8] To control the scale effect of the eigenvalues, we include $\tau_1^{\kappa_2}$ in the γ_{2NT}. Here we implicitly assume that there is at lease one factor.

We compare our agLasso method with the methods of determining the number of factors proposed in Bai and Ng (2002), Onatski (2009, 2010), and Ahn and Horenstein (2013). Their methods only apply to pure factor structures without regressors. Thus, we have to modify their methods to account for the presence of regressors. Specifically, we apply their methods to the factor component: $\lambda_i^{0\prime} F_t^0 + \varepsilon_{it}$, which can be consistently estimated by $\hat{Y}_{it} \equiv Y_{it} - \tilde{\beta}^{c\prime} X_{it}$, where $\tilde{\beta}^c$ is Moon and Weidner's (2014b) bias-corrected estimator of β using the largest number of factors R. We briefly describe their methods here. Bai and Ng (2002, p.201) consider the following information criteria to select the number of factors:

$$
\begin{aligned}
PC_1(r) &= V(r) + r \cdot V(R) \cdot \left(\frac{N+T}{NT}\right) \cdot \ln\left(\frac{NT}{N+T}\right), \\
PC_2(r) &= V(r) + r \cdot V(R) \cdot \left(\frac{N+T}{NT}\right) \cdot \ln\left(\min(N,T)\right), \\
IC_1(r) &= \ln\left(V(r)\right) + r \cdot \left(\frac{N+T}{NT}\right) \cdot \ln\left(\frac{NT}{N+T}\right), \\
IC_2(r) &= \ln\left(V(r)\right) + r \cdot \left(\frac{N+T}{NT}\right) \cdot \ln\left(\min(N,T)\right),
\end{aligned}
$$

where $V(r) = (NT)^{-1} \sum_{i=1}^{N} \hat{\varepsilon}_{i,(r)}' \hat{\varepsilon}_{i,(r)}$, and $\hat{\varepsilon}_{i,(r)}$ is the $T \times 1$ residual vector when r factors are included in the model.

Onatski (2009) develops a test to test the null hypothesis that the true number of factors $R_0 = r$, against the alternative $r < R_0 \leq R$. The test can be used to determine the number of factors. Specifically, we start by testing $H_0 : R_0 = 0$ versus $H_1 : 0 < R_0 \leq R$. If H_0 is not rejected, then we conclude $R_0 = 0$. Otherwise, we continue to test $H_0 : R_0 = 1$ versus $H_1 : 1 < R_0 \leq R$. We repeat the procedure until $H_0 : R_0 = r$ is not rejected and conclude $R_0 = r$. The test is based on the largest eigenvalue of the smoothed periodogram estimate of the spectral density matrix of data and the details are described in Section 4 in Onatski (2009, p. 1455). Onatski (2010) develops an estimator for the number of factors based on the fact that all the "systematic" eigenvalues diverge to infinity.

Ahn and Horenstein (2013) propose the ER (eigenvalue ratio) and GR (growth ratio) estimators for determining the number of factors. The ER estimator maximizes the ratio of two adjacent eigenvalues, while GR estimator maximizes the growth rates of residual variances.

4.3 Effects of the Number of Factors on the Estimation of β's

Before we compare various methods, we first examine the effects of the number of factors included in the model on the performance of the estimators of β's. Table 1 presents the mean squared errors (MSEs) of Moon and Weidner's (2014b) bias-corrected estimators of β_1 and β_2 with different numbers of factors r: $r = 0$, 1, 2, 4, 6 and 8.[9] It is easy to see that when $r = 2$ (the true number of factors), the MSEs are the smallest. In general, the number of factors has substantial effects on the MSEs, especially when N or T is small. For example, when $N = 20$ and $T = 20$ in DGP 3, the MSEs of the estimate of β_1 with $r = 0$ and $r = 8$ are 7 and 3 times as large as those with $r = 2$, respectively; the MSEs of the estimate of β_2 with $r = 0$ and $r = 8$ are 3 and 33 times as large as those with $r = 2$, respectively. The simulations

[9] The results for $r = 3$, 5, and 7 are available upon request. In DGPs 4, 5 and 6 all the regressors are included in the models and the estimation results for the slope coefficients of other regressors are also available upon request.

suggest that the finite sample performance of the estimates of β's crucially depends on the number of included unobservable factors, especially when N or T is small.

4.4 Illustration of the Main Ideas

The main innovation of this paper is to use agLasso to determine the number of factors. There are three key ideas underlying our approach. First, the smallest $R - R_0$ eigenvalues of $\hat{\Sigma}_{\hat{F}}$ $(\equiv \hat{F}'\hat{F}/T)$ converge to zero in probability, while the largest R_0 eigenvalues converge to some positive numbers, which ensures the adaptive nature of our approach. Second, the penalty term λ_{2NT} controls the number of factors selected. The larger the penalty term λ_{2NT} is, the fewer factors are selected. Third, the information criterion chooses an appropriate penalty term λ_{2NT}. Below, we set $R = 8$ and use the simple DGP 1 to illustrate these three main ideas. Note that in DGP 1, there is no regressor so that we only consider the selection of the number of factors.

Plots (a) and (b) in Figure 1 show the medians of the eight eigenvalues of $\hat{\Sigma}_{\hat{F}}$ over the 250 replications for $(N, T) = (20, 20)$ and $(40, 40)$, respectively. It is clear that the two largest eigenvalues are greater than zero, while the six smallest eigenvalues are all close to zero. Plots (c) and (d) show the effects of the penalty term λ_{2NT} $(= C_\gamma s_Y^2 \tau_1^{\kappa_2} \frac{1}{\sqrt{NT}\ln(NT)})$ on the selected number of factors for $(N, T) = (20, 20)$ and $(40, 40)$, respectively. To make the picture clearer, we choose a wide range of C_γ values: 250 points that increase geometrically from 0.001 to 25. We can see that when C_γ (and thus λ_{2NT}) increases, i.e., the penalty becomes larger, a smaller number of factors are selected. We also note that for this DGP, there is quite a wide range of C_γ values (and thus λ_{2NT}) that correctly select the number of factors, especially for $(N, T) = (40, 40)$. Plots (e) and (f) show how our information criterion (IC) changes with respect to C_γ for $(N, T) = (20, 20)$ and $(40, 40)$, respectively. In general, the minimizer of IC falls in the range of C_γ that correctly selects the number of factors.

4.5 Simulation Results

The simulation results are summarized in Tables 2-4. Table 2 reports the proportions of the replications in which the number of factors is correctly determined out of total 250 replications. Our agLasso method is among the best performers in general. For example, for DGP 2, when $N = 20$, $T = 20$, our agLasso method based on our IC selects the true number of factors with a correct rate of 54%, while Bai and Ng's (2002) PC_1, PC_2, IC_1, IC_2 with 0%, 0%, 14%, 8%, respectively, Onatski (2009) with 1%, Onatski (2010) with 22%, and ER and GR estimators of Ahn and Horenstein (2013) with 23% and 25%, respectively. Our "rule of thumb" method for the choice of tuning parameters also gives good results. It has a correct rate of 37%, 49% and 47% for $c = 0.5$, 1 and 2, respectively. When N and T increase, the performances of all the methods improve. In general, among other methods, Bai and Ng's (2002) IC_1 and IC_2, Onatski (2010), and Ahn and Horenstein (2013) are preferred.

Table 3 shows the proportions of the replications in which the estimates of the β's are shrunk to zero out of total 250 replications for DGPs 4-6.[10] Note that all other methods discussed above cannot

[10] To save space, we only report the results based on our IC. The results based on our "rule of thumb" are similar and available upon request. Also for DGP 6, we only report the results for the coefficients of the first five regressors. The results

select regressors. Thus we only report the results using our agLasso method. For DGPs 4 and 6, the relevant regressors $X_{it,1}$ and $X_{it,2}$ are always selected, while a large proportion (e.g., 100% for $N = 60$ and $T = 60$) of the estimated coefficients of the irrelevant regressors (β_3, β_4 and β_5) are shrunk to zero. For DGP 5, when T is large, the relevant regressors are always selected, though a small proportion of relevant regressors are not selected when T is small. For the irrelevant regressors, a large proportion of estimated β_3's and β_4's are shrunk to zero. However, a large proportion of estimated β_5's are shrunk to zero only when T is large.

Table 4 reports the MSEs of estimated β_1's and β_2's for DGPs 2-6 using different methods.[11] In addition to our agLasso estimators, we also report the performances of the bias-corrected agLasso (BC-agLasso) estimator introduced in Section 3.3 and the post-adaptive group Lasso (post-agLasso) estimators, which is Moon and Weidner's (2014b) bias-corrected estimators using the number of factors and regressors selected by our agLasso method. For most of the cases, our agLasso, BC-agLasso, post-agLasso estimators achieve smaller MSEs than other methods when N and T are small. For example, for β_1 in DGP 5 and $(N, T) = (20, 20)$, the 100×MSEs for our agLasso, BC-agLasso and post-agLasso estimators are 4.82, 4.70 and 4.70 respectively, while those for PC_1, PC_2, IC_1, IC_2 in Bai and Ng (2002), Onatski (2009), Onatski (2010), ER and GR in Ahn and Horenstein (2013) are 18.05, 17.89, 17.90, 15.25, 15.51, 8.59, 6.60 and 7.13, respectively. When N and T are large, all the methods perform similarly well. In general, the post-agLasso estimator performs best among the three agLasso-type of estimators, whereas the agLasso and BC-agLasso estimators perform similarly.[12]

5 Empirical Application

In this section, we apply our method to study the determinants of economic growth. There is a large literature on the empirical studies of economic growth. For example, Barro (1991) and Sala-i-Martin et al. (2004) investigate this question using cross-sectional data. For panel data, Islam (1995) employs country fixed effect models and Moral-Benito (2012) uses a Bayesian model averaging approach. Durlauf et al. (2005, DJT) provide a comprehensive literature review. To the best of our knowledge, none of the existing studies allows interactive fixed effects.[13] However, it is plausible that the economic growth is determined not only by observable regressors but also some common unobservable shocks or factors. Thus the panel data model with interactive fixed effects provides much more flexibility in this context. Nevertheless, in practice, we do not have *a priori* knowledge about the number of unobservable factors that should be included in the model. In addition, there are a large number of potential observable variables that may determine economic growth and economic theory does not provide much guidance for the selection of them. For example, DJT survey 145 possible determinants of economic growth and

for the coefficients of the remaining $(K-5)$ irrelevant regressors are available upon request.

[11] For DGPs 4-6, our agLasso method selects the regressors and determines the number of factors, while other method are capable of the latter only. So all the regressors are included for their methods. The estimation results for the coefficients of other $(K-2)$ regressors are available upon request.

[12] For the BC-agLasso estimators, we ignore the fact that the exogenous regessors share the same factor structure as the dependent variable in which case there is no need to correct some of the bias terms ($\mathbb{B}_{2NT}$ in particular). See Remark 6.

[13] The only exception is Su et al. (2015), who apply a specification test of panel models with fixed effects to economic growth data. However, they do not provide estimation results.

point out that "approximately as many growth determinants have been proposed as there are countries for which data are available. It is hard to believe that all these determinants are central...". Thus, it is important to determine the number of factors and select the relevant regressors in this context.

5.1 Data and Implementation

Let Y_{it} be the growth rate of real GDP per capita for country i in year t. X_{it} includes 9 variables as listed in Table 5. The sample covers 108 countries for the period 1970-2005, i.e., $N = 108$ and $T = 36$. Here we need a balanced panel, and the selection of the 108 countries is completely determined by the availability of data.[14] The data sources are the Penn World Table (Penn Table) and World Bank World Development Indicators (WDI).

We include maximum 8 factors in the model. As in the simulations, we choose $(\gamma_{1NT}, \gamma_{2NT})$ from the set $\left\{C_\gamma s_Y^2 \frac{1}{\sqrt{NT}\ln(NT)}\right\} \times \left\{C_\gamma s_Y^2 \tau_1^{\kappa_2} \frac{1}{\sqrt{NT}\ln(NT)}\right\}$, where C_γ are 100 constants that increase geometrically from 0.01 to 25. Both κ_1 and κ_2 are equal to 2. The information criterion is the same as that in the simulations.

We also consider other methods, including IC_1 and IC_2 in Bai and Ng (2002) and the methods in Onatski (2010) and Ahn and Horenstein (2013), as our simulations show that they are preferred methods. In the case that different methods give conflicting conclusions, we can use a simple majority rule.

5.2 Estimation Results

5.2.1 Estimation without regressors

We first consider a pure factor structure without including any regressors. Our agLasso method chooses **3** factors. The eigenvalues $(\tau_1, ..., \tau_8)$ used in our agLasso are shown in Figure 2(a). Other methods also choose **3** factors as shown in Table 6.

5.2.2 Linear estimation

We consider a linear specification that uses the 9 variables listed in Table 5 and the first 3 lags of Y_{it} as regressors. The first half of Table 6 shows the estimation results for the different numbers of factors, $r = 0$, 3, 5 and 8.[15] The estimates of the coefficients vary substantially with different numbers of factors. For example, the coefficient of consumption is negative and significant when r is smaller than 4 and becomes insignificant when r is greater than 4. However, the coefficients of government consumption share, investment share, and the first lag of economic growth are significant for most of the numbers of factors.

Our agLasso method chooses **3** factors, which is consistent with Bai and Ng's (2002) IC_1 and IC_2 and Onatski's (2010) method as shown in Table 6. Ahn and Horenstein (2013) choose 1 factor. The eigenvalues used in our agLasso are shown in Figure 2(b). The estimation results are presented in Table 7. Our agLasso selects five regressors: population growth, government consumption share, investment share, and the first and second lags of economic growth. Among them, government consumption share,

[14] The list of the 108 countries is available upon request.

[15] The results for the number of factors $r = 1$, 2, 4, 6 and 7 are available upon request.

investment share, and the first lag of economic growth are significant. The government consumption share has a negative effect on economic growth, while the investment share and lagged economic growth have positive effects.

5.2.3 Nonlinear estimation

In this subsection, we examine the nonlinear effects of the regressors. Our agLasso method selects five regressors in the linear specification. Thus, we include the squared and interaction terms of those five selected regressors in addition to the 12 regressors in the linear specification. The total number of regressors included is 27. In this case, all methods except Ahn and Horenstein (2013) select **3** factors again. The eigenvalues used in our agLasso are shown in Figure 2(c). Among the 27 regressors, our agLasso method selects **9** regressors. Table 8 presents the estimation results for the 9 selected regressors. Based on the post-agLasso method, we find that consumption share, investment share, and the interaction term of government consumption share and investment share are significant at the 1% level, while the first lag of economic growth is significant at the 10% level. The signs of those significant regressors are in general consistent with those in the linear specification. The government consumption share has a negative effect through its interaction term with investment share, while the lagged economic growth has positive effects. The effect of investment share is $0.219 - 0.008 \times \text{Gov}$, which is positive for most values of government consumption share in the sample. However, population growth becomes insignificant, and consumption share becomes significant with a negative sign.

To further examine the nonlinear effect, we consider a "high dimensional" model by including the linear, squared, and interaction terms of all the original 12 regressor (i.e., the 9 variables listed in Table 5 and the first 3 lags of the economic growth). The total number of regressors is 90 in this case. All methods select **3** factors again. Among the 90 regressors, our agLasso method selects **11** regressors as shown in Table 9. The first half of Table 8 also reports the estimation results for the 11 regressors using the model that includes all the 90 regressors with different numbers of factors.[16] Note that almost all the regressors are insignificant when all 90 regressors are included. This is not surprising, as the standard errors can be easily inflated when a large number of regressors are included. Nevertheless, our agLasso is effective in selecting the relevant regressors. Based on the post-agLasso estimation results, consumption share, government consumption share, investment share, the first lag of economic growth and the interaction term of fertility rates and the lagged economic growth are significant. Among them, consumption share and government consumption share have negative effects, while investment share and lagged economic growth have positive effects on economic growth.[17]

To summarize, we find that in general there are 3 unobservable factors that determine economic growth. Among the observable regressors, considering both linear and nonlinear specifications, we find that government consumption share has a negative effect, while investment share and lagged economic growth have positive effects on economic growth. This finding is largely consistent with the existing empirical literature on economic growth (see, e.g., DJT, Appendix 2).

[16] The estimation results for the other 79 regressors are available upon request.

[17] The effect of the first lag of economic growth is $0.321 - 0.049 \times \text{Fert}$, which is positive in general.

6 Conclusion

We propose a novel adaptive group Lasso procedure for simultaneous selection of factors and relevant regressors and estimation in dynamic panel models with interactive fixed effects. We show that our method consistently determines the number of factors and selects relevant regressors. Our estimators of the slope parameters in the models achieve an oracle property. Our simulations suggest that our new method performs well in finite samples. We apply our method to study the determinants of economic growth and find that government consumption share has negative effects, whereas investment share and lagged economic growth have positive effects on economic growth.

There are several interesting topics for further research. First, we only allow the numbers of relevant regressors (K_0) to grow with the sample size but assume that the true number of factors (R_0) is fixed in this paper. The divergence of K_0 to infinity is particularly useful for the nonparametric sieve estimation of dynamic panel models with interactive fixed effects (see, e.g., Su and Zhang (2014)). But it is also desirable to extend our method to allow R_0 to increase with both N and T. Second, we only consider strong factors in our model. As a referee points out, it is interesting to focus on the pure factor model with weak or semi-strong factors and compare our method with Bai and Ng's (2002) method in the determination of the number of factors. To fix the idea, we can assume that the factors are well normalized such that $T^{-1}F'F$ has a well-behaved probability limit but allow

$$N^{-a_r} \|\lambda_{\cdot r}\| \xrightarrow{P} c_r \in (0, \infty) \text{ for } r = 1, ..., R,$$

where $a_r \in [0, 1/2]$ for $r = 1, ..., R$. Apparently, $a_r = 0$ and $1/2$ correspond to the weak factors studied in Onatski (2012) and the commonly studied strong factors, respectively. Preliminary simulations indicate the good performance of our approach in comparison with Bai and Ng's (2002) when the factors are semi-strong (e.g., $a_r = 1/4$). We conjecture that by allowing for different degrees of strength for different factors, our shrinkage method can also help identify strong or semi-strong factors and separate them from those relatively weak or inessential factors, but we leave the rigorous theoretic analysis for future research. Third, as an alternative to the adaptive group Lasso used in this paper, the SCAD method of Fan and Li (2001) can also be employed and it will be interesting to compare our method with that based on the SCAD method. Fourth, it is also possible to allow endogeneity in panel data models with interactive fixed effects, in which case various important issues would arise, including how to extend the usual instrumental variable (IV) estimation to the current framework, how to determine the set of instruments, and how to select the number of factors and relevant regressors. Endogeneity naturally arises in dynamic panel data models with measurement error (e.g., Lee et al., 2012) and in various macro and micro panel data models (e.g., Moon et al., 2014). Fifth, instead of considering variable and factor selection in augmented panel regression models, an alternative is to consider SCAD- or Bridge-based penalized PCA in one step. We are exploring some of these topics in ongoing works.

APPENDIX

A Some Technical Lemmas

Recall $\hat{\mathbf{Y}} = \mathbf{Y} - \sum_{k=1}^{K} \tilde{\beta}_k^c \mathbf{X}_k$, $\hat{F} = (NT)^{-1} \hat{\mathbf{Y}}' \hat{\mathbf{Y}} \tilde{F}$, $H = H_{NT} = \left(N^{-1}\lambda^{0\prime}\lambda^0\right)(T^{-1}F^{0\prime}\tilde{F})$, and $T^{-1}\tilde{F}'\tilde{F} = I_R$. Recall $\tilde{F} = (\tilde{F}_{(1)}, \tilde{F}_{(2)})$, $\hat{F} = (\hat{F}_{(1)}, \hat{F}_{(2)})$ and $H = (H_{(1)}, H_{(2)})$, where, e.g., $\tilde{F}_{(1)}$ and $\tilde{F}_{(2)}$ are $T \times R_0$ and $T \times (R - R_0)$ submatrices, respectively. Noting that $T^{-1}||\tilde{F}||^2 = \mathrm{tr}(T^{-1}\tilde{F}'\tilde{F}) = R$, $||\tilde{F}|| = O_P\left(T^{1/2}\right)$. Write $V_{NT} = \mathrm{diag}(V_{NT,11},\ V_{NT,22})$ where $V_{NT,11}$ and $V_{NT,22}$ are $R_0 \times R_0$ and $(R - R_0) \times (R - R_0)$ submatrices of V_{NT}, respectively. Let $F_{(1)}^* = F^0 H_{(1)}$ and $\lambda_{(1)}^* = \lambda^0 H_{(1)}^{+\prime}$. For matrices G and $\tilde{G}$, we write $\tilde{G} = G + \mathbf{O}_P(c_{NT})$ if $\left\|\tilde{G} - G\right\| = O_P(c_{NT})$ and $\tilde{G} = G + \mathbf{o}_P(c_{NT})$ if $\left\|\tilde{G} - G\right\| = o_P(c_{NT})$. Note that $\mathbf{O}_P$ and $\mathbf{o}_P$ are equivalent to O_P and o_P respectively when the associated matrices are of finite (fixed) dimensions.

We first state some technical lemmas whose proofs are given in the supplementary Appendix C.

Lemma A.1 *Suppose that Assumptions A.1 and A.3(i) hold. Then*

(i) $T^{-1}\tilde{F}'(NT)^{-1}\hat{\mathbf{Y}}'\hat{\mathbf{Y}}\tilde{F} = V_{NT} \xrightarrow{P} V$,

(ii) $(T^{-1}\tilde{F}'F^0)\left(N^{-1}\lambda^{0\prime}\lambda^0\right)(T^{-1}F^{0\prime}\tilde{F}) \xrightarrow{P} V$,

(iii) $T^{-1}F^{0\prime}\tilde{F}_{(1)} \xrightarrow{P} \Delta_1$ *and* $T^{-1}F^{0\prime}\tilde{F}_{(2)} \xrightarrow{P} 0$,

(iv) $H_{(1)} \xrightarrow{P} H_{(1)}^0 = \Sigma_{\lambda^0}\Delta_1$ *and* $H_{(2)} \xrightarrow{P} 0$,

where V_{NT} *is an* $R \times R$ *diagonal matrix consisting of the* R *largest eigenvalues of* $(NT)^{-1}\hat{\mathbf{Y}}'\hat{\mathbf{Y}}$, *and* $V = \begin{bmatrix} V_{11} & 0 \\ 0 & 0 \end{bmatrix}$ *with* V_{11} *being an* $R_0 \times R_0$ *matrix consisting of the* R_0 *eigenvalues of* $\Sigma_{\lambda^0}\Sigma_{F^0}$, *both arranged in descending order;* Δ_1 *is an* $R_0 \times R_0$ *full rank matrix.*

Lemma A.2 *Suppose that Assumptions A.1, A.2 and A.3(i) hold. Let* $\breve{\tau}_1, ..., \breve{\tau}_R$ *denote the eigenvalues of* $T^{-1}H'F^{0\prime}F^0H$ *in descending order. Then*

(i) $T^{-1}\left\|\hat{F} - F^0H\right\|^2 = O_P(\delta_{NT}^{-2})$,

(ii) $T^{-1}\left\|\tilde{F}_{(1)} - F^0H_{(1)}V_{NT,11}^{-1}\right\|^2 = O_P(\delta_{NT}^{-2})$ *and* $\left\|H_{(2)}\right\| = O_P(\delta_{NT}^{-1})$,

(iii) $T^{-1}(\hat{F}_{(1)} - F^0H_{(1)})'F^0 = O_P(\delta_{NT}^{-2} + (NT/K)^{-1/2})$ *and* $T^{-1}(\hat{F}_{(2)} - F^0H_{(2)})'F^0 = O_P(N^{-1/2})$,

(iv) $T^{-1}(\hat{F}_{(1)} - F^0H_{(1)})'\hat{F} = O_P(\delta_{NT}^{-2} + (NT/K)^{-1/2})$, $T^{-1}(\hat{F}_{(2)} - F^0H_{(2)})'\hat{F} = O_P(N^{-1/2})$, *and* $T^{-1}(\hat{F} - F^0H)'\hat{F}_{(2)} = O_P(\delta_{NT}^{-2})$,

(v) $T^{-1}(\hat{F}_{(1)}'\hat{F}_{(1)} - H_{(1)}'F^{0\prime}F^0H_{(1)}) = O_P(\delta_{NT}^{-2} + (NT/K)^{-1/2})$ *and* $T^{-1}(\hat{F}'\hat{F} - H'F^{0\prime}F^0H) = O_P(N^{-1/2})$,

(vi) $\left\|P_{\hat{F}_{(1)}} - P_{F_{(1)}^*}\right\| = O_P(\delta_{NT}^{-1})$,

(vii) $\tau_l - \breve{\tau}_l = O_P(N^{-1/2})$ *for* $l = 1, 2, ..., R_0$ *and* $\tau_l = O_P(N^{-1/2})$ *for* $l = R_0 + 1, ..., R_0$, *where* $\tau_1, ..., \tau_R$ *denote the* R *eigenvalues of* $T^{-1}\hat{F}'\hat{F}$ *arranged in descending order.*

Lemma A.3 *Suppose that Assumptions A.1, A.2, A.3(i), and A.6(i) hold. Let* $\lambda_{i(1)}^{*\prime}$ *denote the ith row of* $\lambda_{(1)}^*$. *Then*

(i) $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} X_{i(1)}'(P_{\hat{F}_{(1)}} - P_{F_{(1)}^*})(\hat{F}_{(1)} - F_{(1)}^*)\lambda_{i(1)}^* = \mathbb{B}_{1NT} + \mathbf{o}_P(1)$,

(ii) $\frac{1}{N}\sum_{i=1}^{N} X_{i(1)}'(\hat{F}_{(1)} - F_{(1)}^*)(\hat{F}_{(1)}'\hat{F}_{(1)})^{-1}F_{(1)}^{*\prime}\varepsilon_i = \mathbf{O}_P(K_0^{1/2}T^{1/2}\delta_{NT}^{-2}) = \mathbf{o}_P(1)$,

(iii) $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} X_{i(1)}'F_{(1)}^*(\hat{F}_{(1)}'\hat{F}_{(1)})^{-1}(\hat{F}_{(1)} - F_{(1)}^*)'\varepsilon_i = \mathbb{B}_{3NT} + \mathbf{O}_P(K_0^{1/2}T^{1/2}\delta_{NT}^{-2}) = \mathbb{B}_{3NT} + \mathbf{o}_P(1)$,

(iv) $\frac{1}{T}(\hat{F}_{(1)} - F_{(1)}^*)'\varepsilon_i = O_P(\delta_{NT}^{-2})$ *for* $i = 1, ..., N$,

where $\mathbb{B}_{1NT} = \frac{1}{N^{5/2}T^{5/2}} \sum_{i=1}^{N} X'_{i(1)} F^*_{(1)} (F^{*\prime}_{(1)} F^*_{(1)})^{-1} \tilde{F}'_{(1)} F^0 \lambda^{0\prime} \boldsymbol{\varepsilon\varepsilon}' \lambda^0 F^{0\prime} \tilde{F}_{(1)} \lambda^0_{i(1)}$, $\mathbb{B}_{3NT} = \sqrt{\frac{T}{N}} b_{3NT}$ *and* $b_{3NT} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} (T^{-1} X'_{i(1)} F^*_{(1)})(T^{-1} F^{*\prime}_{(1)} F^*_{(1)})^{-1} \ (T^{-1} \tilde{F}'_{(1)} F^0) \lambda^0_j (T^{-1} \varepsilon'_j \varepsilon_i)$.

B Proofs of the Main Results

In this appendix, we prove the main results in the paper.

B.1 Proof of Theorem 3.1

The proof is done in the same spirit of Fan and Li (2001), Fan and Peng (2004), and Lam and Fan (2008). In particular, the latter two papers consider estimation with a diverging number of parameters. Recall that $\lambda^*_i = H^+ \lambda^0_i$ and $\lambda^* = (\lambda^*_1, ..., \lambda^*_N)' = \lambda^0 H^{+\prime}$. Let $a_{NT} = T^{-1/2}$. Let $\beta = \beta^0 + a_{NT} b$ and $\lambda = \lambda^* + a_{NT} u$, where $b = (b_1, ..., b_K)'$ and $u = (u_{\cdot 1}, ..., u_{\cdot R})$ are matrices of dimensions $K \times 1$ and $N \times R$, respectively. Apparently, we use $u_{\cdot r}$ to denote the rth column of u for $r = 1, 2, ..., R$. Let u'_i denote the ith row of $u : u = (u_1, ..., u_N)'$. Our aim is to show that for any given $\epsilon > 0$, there exists a large constant L such that for sufficiently large (N, T) we have

$$P\left\{ \inf_{\|b\|=L,\ N^{-1/2}\|uH'\|=L} Q_\gamma\left(\beta^0 + a_{NT} b, \lambda^* + a_{NT} u\right) > Q_\gamma\left(\beta^0, \lambda^*\right) \right\} \geq 1 - \epsilon. \tag{B.1}$$

This implies that with probability approaching one (w.p.a.1) there is a local minimum $(\hat\beta, \hat\lambda)$ such that either $\hat\beta$ lie inside the ball $\left\{ (\beta^0 + a_{NT} b) : \|b\| \leq L \right\}$, or $\hat\lambda$ lies inside the ball $\left\{ (\lambda^* + a_{NT} u) : N^{-1/2} \|u\| \leq L \right\}$, or both. Then we have $\left\| \hat\beta - \beta^0 \right\| = O_P(a_{NT})$, or $N^{-1/2} \left\| (\hat\lambda - \lambda^*) H' \right\| = O_P(a_{NT})$, or both.

Let $D_\gamma(b, u) \equiv Q_\gamma\left(\beta^0 + a_{NT} b, \lambda^* + a_{NT} u\right) - Q_\gamma\left(\beta^0, \lambda^*\right)$. By (2.6),

$$\begin{aligned}
D_\gamma(b, u) &= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\left(Y_{it} - \beta^{0\prime} X_{it} - \lambda^{*\prime}_i \hat{F}_t - a_{NT} b' X_{it} - a_{NT} u'_i \hat{F}_t \right)^2 - \left(Y_{it} - \beta^{0\prime} X_{it} - \lambda^{*\prime}_i \hat{F}_t \right)^2 \right] \\
&\quad + \gamma_{1NT} \sum_{k=1}^{K} \frac{1}{\left| \tilde\beta^c_k \right|^{\kappa_1}} \left(\left| \beta^0_k + a_{NT} b_k \right| - \left| \beta^0_k \right| \right) + \frac{\gamma_{2NT}}{\sqrt{N}} \sum_{r=1}^{R} \frac{1}{|\tau_r|^{\kappa_2}} \left(\|\lambda^*_{\cdot r} + a_{NT} u_{\cdot r}\| - \|\lambda^*_{\cdot r}\| \right) \\
&= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(a_{NT} b' X_{it} + a_{NT} u'_i H' F^0_t \right)^2 - \frac{2}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{it} \left(a_{NT} b' X_{it} + a_{NT} u'_i H' F^0_t \right) \\
&\quad + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\left(a_{NT} b' X_{it} + a_{NT} u'_i \hat{F}_t \right)^2 - \left(a_{NT} b' X_{it} + a_{NT} u'_i H' F^0_t \right)^2 \right] \\
&\quad - \frac{2}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\breve\varepsilon_{it} \left(a_{NT} b' X_{it} + a_{NT} u'_i \hat{F}_t \right) - \varepsilon_{it} \left(a_{NT} b' X_{it} + a_{NT} u'_i H' F^0_t \right) \right] \\
&\quad + \gamma_{1NT} \sum_{k=1}^{K} \frac{1}{\left| \tilde\beta^c_k \right|^{\kappa_1}} \left(\left| \beta^0_k + a_{NT} b_k \right| - \left| \beta^0_k \right| \right) + \frac{\gamma_{2NT}}{\sqrt{N}} \sum_{r=1}^{R} \frac{1}{|\tau_r|^{\kappa_2}} \left(\|\lambda^*_{\cdot r} + a_{NT} u_{\cdot r}\| - \|\lambda^*_{\cdot r}\| \right) \\
&\equiv A_1(b, u) - 2A_2(b, u) + A_3(b, u) - 2A_4(b, u) + A_5(b) + A_6(u), \text{ say,}
\end{aligned}$$

where $\breve\varepsilon_{it} = Y_{it} - \beta^{0\prime} X_{it} - \lambda^{*\prime}_i \hat{F}_t$. We want to determine the probability order of A_j's. First,

$$A_1(b, u) = a^2_{NT} b' A b + a^2_{NT} \text{tr}\left(\frac{uH'}{\sqrt{N}} \frac{F^{0\prime} F^0}{T} \frac{Hu'}{\sqrt{N}} \right) + 2a^2_{NT} \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \text{tr}\left(\frac{uH'}{\sqrt{N}} F^0_t b' X'_{\cdot t} \right),$$

where $A = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}X_{it}' = \bar{W}_{NT}$, and $X_{\cdot t}$ is an $N \times K$ matrix with the ith row given by X_{it}'. Second, $A_2(b,u) = a_{NT}b'\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}X_{it} + \frac{a_{NT}}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}F_t^{0\prime}Hu_i \equiv A_{2,1}(b) + A_{2,2}(u)$, say. Noting that $\left\|(NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}X_{it}\right\| = O_P(K^{1/2})$ by Assumption A.1(vi), $A_{2,1}(b) = O_P(a_{NT}$ $(NT/K)^{-1/2})\|b\| = o_P(a_{NT}^2)\|b\|$. By Cauchy-Schwarz (CS hereafter) inequality, Assumption A.1(viii), and the fact that $\|H\| = O_P(1)$, we have $|A_{2,2}(u)| = \frac{a_{NT}}{NT}|\mathrm{tr}(\boldsymbol{\varepsilon}F^0Hu')| \leq \frac{a_{NT}}{NT}\left\|\boldsymbol{\varepsilon}F^0\right\|\|H\|\|u\| = O_P(a_{NT}^2N^{-1/2})\|u\|$. It follows that $A_2(b,u) = o_P(a_{NT}^2)\|b\| + O_P(a_{NT}^2N^{-1/2})\|u\|$. Third, using the fact that $a_1^2 - a_2^2 = (a_1-a_2)^2 + 2(a_1-a_2)a_2$, we have

$$\begin{aligned}
A_3(b,u) &= \frac{a_{NT}^2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_i'\left(\hat{F}_t - H'F_t^0\right)\left(\hat{F}_t - H'F_t^0\right)'u_i + \frac{2a_{NT}^2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_i'\left(\hat{F}_t - H'F_t^0\right)F_t^0H'u_i \\
&\quad + \frac{2a_{NT}^2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}b'X_{it}\left(\hat{F}_t - H'F_t^0\right)'u_i \\
&\equiv A_{3,1}(u) + 2A_{3,2}(u) + 2A_{3,3}(b,u), \text{ say.}
\end{aligned}$$

By the submultiplicative property of the Frobenius norm, CS inequality, and Lemmas A.2(i) and (iii),

$$\begin{aligned}
|A_{3,1}(u)| &\leq \frac{a_{NT}^2}{N}\frac{1}{T}||\hat{F} - F^0H||^2\|u\|^2 = O_P\left(a_{NT}^2N^{-1}\delta_{NT}^{-2}\right)\|u\|^2 = o_P\left(a_{NT}^2N^{-1}\right)\|u\|^2, \\
|A_{3,2}(u)| &\leq \frac{a_{NT}^2}{N}\frac{1}{T}||(\hat{F} - F^0H)F^0H||\,\|u\|^2 = O_P\left(a_{NT}^2N^{-3/2}\right)\|u\|^2 = o_P\left(a_{NT}^2N^{-1}\right)\|u\|^2,
\end{aligned}$$

and

$$\begin{aligned}
|A_{3,3}(b,u)| &\leq \frac{a_{NT}^2}{(NT/K)^{1/2}}T^{-1/2}||\hat{F} - F^0H||\left\{\frac{1}{NK}\sum_{i=1}^{N}\|X_i\|^2\right\}^{1/2}\|b\|\|u\| \\
&= O_P(a_{NT}^2(NT/K)^{-1/2}\delta_{NT}^{-1})\|b\|\|u\| = o_P(a_{NT}^2N^{-1/2})\|b\|\|u\| \\
&\leq o_P(a_{NT}^2)\|b\|^2 + o_P\left(a_{NT}^2N^{-1}\right)\|u\|^2.
\end{aligned}$$

Thus $A_3(b,u) = o_P(a_{NT}^2)\|b\|^2 + o_P\left(a_{NT}^2N^{-1}\right)\|u\|^2$.

Next, let $\breve{\varepsilon}_i \equiv (\breve{\varepsilon}_{i1},...,\breve{\varepsilon}_{iT})'$. In view of the fact that $\breve{\varepsilon}_{it} - \varepsilon_{it} = -\lambda_i^{*\prime}(\hat{F}_t - H'F_t^0)$, we have $\breve{\varepsilon}_i - \varepsilon_i = -(\hat{F} - F^0H)\lambda_i^*$ and

$$\begin{aligned}
A_4(b,u) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\breve{\varepsilon}_{it}\left(a_{NT}b'X_{it} + a_{NT}u_i'\hat{F}_t\right) - \varepsilon_{it}\left(a_{NT}b'X_{it} + a_{NT}u_i'H'F_t^0\right)\right] \\
&= b'\frac{a_{NT}}{NT}\sum_{i=1}^{N}X_i'(\breve{\varepsilon}_i - \varepsilon_i) + \frac{a_{NT}}{NT}\sum_{i=1}^{N}(\breve{\varepsilon}_i - \varepsilon_i)'\hat{F}u_i + \frac{a_{NT}}{NT}\sum_{i=1}^{N}\varepsilon_i'(\hat{F} - F^0H)u_i \\
&= -b'\frac{a_{NT}}{NT}\sum_{i=1}^{N}X_i'(\hat{F} - F^0H)\lambda_i^* - \frac{a_{NT}}{NT}\sum_{i=1}^{N}\lambda_i^{*\prime}(\hat{F} - F^0H)\hat{F}u_i + \frac{a_{NT}}{NT}\sum_{i=1}^{N}\varepsilon_i'(\hat{F} - F^0H)u_i \\
&\equiv -A_{4,1}(b) - A_{4,2}(u) + A_{4,3}(u), \text{ say.}
\end{aligned}$$

Using the fact that $\lambda_i^* = H^+\lambda_i^0 = H'(HH')^{-1}\lambda_i^0$ and $H = (H_{(1)},\, H_{(2)})$, we have

$$\begin{aligned} A_{4,1}(b) &= b'\frac{a_{NT}}{NT}\sum_{i=1}^{N} X_i'(\hat{F}_{(1)} - F^0H_{(1)})H_{(1)}'(HH')^{-1}\lambda_i^0 \\ &\quad + b'\frac{a_{NT}}{NT}\sum_{i=1}^{N} X_i'(\hat{F}_{(2)} - F^0H_{(2)})H_{(2)}'(HH')^{-1}\lambda_i^0 \\ &\equiv A_{4,1,1}(b) + A_{4,1,2}(b), \text{ say.} \end{aligned}$$

Following the proof of Lemma A.3(iii) in Bai (2009), we can show that $\frac{1}{NT}\sum_{i=1}^{N} X_i'(\hat{F}_{(1)} - F^0H_{(1)})H_{(1)}' \times (HH')^{-1}\lambda_i^0 = O_P\left(K^{1/2}\delta_{NT}^{-2}\right)$. It follows that $A_{4,1,1}(b) = O_P\left(a_{NT}K^{1/2}\delta_{NT}^{-2}\right)\|b\| = o_P\left(a_{NT}^2\right)\|b\|$ by Assumption A.3(i). For $A_{4,1,2}(b)$, we have by the triangle and CS inequalities, and Lemmas A.2(i)-(ii),

$$\begin{aligned} |A_{4,1,2}(b)| &\leq K^{1/2}a_{NT}\left\{\frac{1}{T^{1/2}}\left\|\hat{F}_{(2)} - F^0H_{(2)}\right\|\right\}\left\{\frac{1}{NTK}\sum_{i=1}^{N}\|X_i\|^2\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i^0\right\|^2\right\}^{1/2} \\ &\quad \times \left\|H_{(2)}\right\| \, ||(HH')^{-1}|| \, \|b\| \\ &= K^{1/2}a_{NT}O_P\left(\delta_{NT}^{-1}\right)O_P(1)\,O_P\left(\delta_{NT}^{-1}\right)O_P(1)\|b\| = o_P\left(a_{NT}^2\right)\|b\|. \end{aligned}$$

Then $A_{4,1}(b) = o_P\left(a_{NT}^2\right)\|b\|$. Note that

$$A_{4,2}(u) = \frac{a_{NT}}{NT}\mathrm{tr}[\lambda^*(\hat{F} - F^0H)'F^0Hu'] + \frac{a_{NT}}{NT}\mathrm{tr}[\lambda^*(\hat{F} - F^0H)'(\hat{F} - F^0H)u'] \equiv A_{4,2,1}(u) + A_{4,2,2}(u).$$

Using $\lambda^* = \lambda^0H^{+\prime} = \lambda^0(HH')^{-1}H$, we further decompose $A_{4,2,1}(u)$ as follows $A_{4,2,1}(u) = \frac{a_{NT}}{NT}\mathrm{tr}[\lambda^0(HH')^{-1} H_{(1)}(\hat{F}_{(1)} - F^0H_{(1)})'F^0Hu'] + \frac{a_{NT}}{NT}\mathrm{tr}[\lambda^0(HH')^{-1}H_{(2)}(\hat{F}_{(2)} - F^0H_{(2)})'F^0Hu'] \equiv A_{4,2,1a}(u) + A_{4,2,1b}(u)$, say. By CS inequality, Lemma A.2(iii), and the fact that $\delta_{NT}^{-2} + K(NT)^{-1/2} = o(T^{-1/2})$ under Assumption A.3(i), we have

$$\begin{aligned} |A_{4,2,1a}(u)| &\leq \frac{a_{NT}}{N^{1/2}}\frac{1}{N^{1/2}}\|\lambda^0\|\left\|(HH')^{-1}H_{(1)}\right\|\frac{1}{T}\left\|(\hat{F}_{(1)} - F^0H_{(1)})'F^0H\right\|\|u\| \\ &= O_P\left(a_{NT}N^{-1/2}(\delta_{NT}^{-2} + K(NT)^{-1/2})\right)\|u\| = o_P\left(a_{NT}^2N^{-1/2}\right)\|u\|. \end{aligned}$$

Similarly, by Lemmas A.2(ii)-(iii), we can show that $A_{4,2,1b}(u) = o_P\left(a_{NT}^2N^{-1/2}\right)\|u\|$. In addition, by Lemma A.2(i), we can readily show that

$$A_{4,2,2}(u) = \frac{a_{NT}}{N^{1/2}}\frac{1}{N^{1/2}}\|\lambda^*\|\frac{1}{T}\left\|\hat{F} - F^0H\right\|^2\|u\| = O_P\left(a_{NT}N^{-1/2}\delta_{NT}^{-2}\right)\|u\| = o_P\left(a_{NT}^2N^{-1/2}\right)\|u\|.$$

It follows that $A_{4,2}(u) = o_P\left(a_{NT}^2N^{-1/2}\right)\|u\|$. By the fact that $|\mathrm{tr}(A_1)| \leq \mathrm{rank}(A_1)\|A_1\|_{\mathrm{sp}}$, $\|A_1A_2\|_{\mathrm{sp}} \leq \|A_1\|_{\mathrm{sp}}\|A_2\|_{\mathrm{sp}}$ and $\|A_1\|_{\mathrm{sp}} \leq \|A_1\|$, Lemma A.2(i), and Assumption A.1(v),

$$\begin{aligned} |A_{4,3}(u)| &= \frac{a_{NT}}{NT}\left|\mathrm{tr}\left(\boldsymbol{\varepsilon}\left(\hat{F} - F^0H\right)u'\right)\right| \leq \frac{Ra_{NT}}{NT^{1/2}}\|\boldsymbol{\varepsilon}\|_{\mathrm{sp}}\frac{1}{T^{1/2}}\left\|\hat{F} - F^0H\right\|\|u\| \\ &= \frac{Ra_{NT}}{NT^{1/2}}O_P\left(T^{1/2} + N^{1/2}\right)O_P\left(\delta_{NT}^{-1}\right)\|u\| = o_P\left(a_{NT}^2N^{-1/2}\right)\|u\|. \end{aligned}$$

Consequently, $A_4(b,u) = o_P\left(a_{NT}^2\right)\|b\| + o_P\left(a_{NT}^2N^{-1/2}\right)\|u\|$. Noting that $A_5(b) \geq \gamma_{1NT}\sum_{k=1}^{K_0}\frac{1}{|\tilde{\beta}_k^c|^{\kappa_1}}(|\beta_k^0 + a_{NT}b_k| - |\beta_k^0|)$, by the triangle and CS inequalities and Assumption A.3(ii), we have

$$\left|\gamma_{1NT}\sum_{k=1}^{K_0}\frac{1}{\left|\tilde{\beta}_k^c\right|^{\kappa_1}}(\left|\beta_k^0 + a_{NT}b_k\right| - \left|\beta_k^0\right|)\right| \leq O_P(a_{NT}\gamma_{1NT})\sum_{k=1}^{K_0}|b_k| \leq O_P(K_0^{1/2}a_{NT}\gamma_{1NT})\|b\| = o_P(a_{NT}^2)\|b\|,$$

we have $A_5(b) \geq -o_P(a_{NT}^2)\|b\|$. Similarly, $A_6(u) \geq -o_P(a_{NT}^2 N^{-1/2})\|u\|$. It follows that

$$D_\gamma(b,u) \geq \Pi_{NT}(b,u) + o_P(a_{NT}^2)\left\{K^{-1}\|b\|^2 + K^{-1/2}\|b\| + N^{-1}\|u\|^2 + N^{-1/2}\|u\|\right\} = \Pi_{NT}(b,u) + \text{s.m.},$$

where

$$\Pi_{NT}(b,u) = a_{NT}^2\left[b'Ab + \operatorname{tr}\left(\frac{uH'}{\sqrt{N}}\frac{F^{0\prime}F^0}{T}\frac{Hu'}{\sqrt{N}}\right) + \frac{2}{N^{1/2}T}\sum_{t=1}^{T}\operatorname{tr}\left(\frac{uH'}{\sqrt{N}}F_t^0 b' X_{\cdot t}'\right) - \frac{2}{N^{1/2}T^{1/2}}\operatorname{tr}\left(\boldsymbol{\varepsilon}F^0\frac{Hu'}{\sqrt{N}}\right)\right],$$

and s.m. denotes terms of smaller order than the leading term. Let $B = \left(\frac{F^{0\prime}F^0}{T}\right)\otimes I_N$, $C = \frac{1}{\sqrt{N}T}\sum_{t=1}^{T}F_t^0\otimes X_{\cdot t}$, $D = \frac{1}{N^{1/2}T^{1/2}}\boldsymbol{\varepsilon}F^0$, and $\eta = \frac{1}{N^{1/2}}\operatorname{vec}(uH')$. Following Bai (2009, p.1265), we have

$$\begin{aligned}
\Pi_{NT}(b,u) &= a_{NT}^2\left[b'Ab + \eta' B\eta + 2b'C'\eta - 2\eta'\operatorname{vec}(D)\right] \\
&= a_{NT}^2\left[b'\left(A - C'B^{-1}C\right)b + \left(\eta' + b'C'B^{-1}\right)B\left(\eta + B^{-1}Cb\right) - 2\eta'\operatorname{vec}(D)\right] \\
&= a_{NT}^2\left[b'\bar{A}b + \bar{\eta}'B\bar{\eta} - 2\left(\bar{\eta} - B^{-1}Cb\right)'\operatorname{vec}(D)\right] \\
&= a_{NT}^2\left[b'\bar{A}b + 2b'C'B^{-1}\operatorname{vec}(D)\right] + a_{NT}^2\left[\bar{\eta}'B\bar{\eta} - 2\bar{\eta}'\operatorname{vec}(D)\right],
\end{aligned}$$

where $\bar{A} = A - C'B^{-1}C$, $\bar{\eta} = \eta + B^{-1}Cb$, and the first equality follows from the fact that $\operatorname{tr}(B_1B_2) = \operatorname{vec}(B_2')' \times\operatorname{vec}(B_1)$, $\operatorname{tr}(B_1B_2B_3) = \operatorname{vec}(B_1)'(B_2\otimes I)\operatorname{vec}(B_3')$, and $\operatorname{tr}(B_1B_2B_3B_4) = \operatorname{vec}(B_1)'(B_2\otimes B_4')\operatorname{vec}(B_3')$ for any conformable matrices B_1, B_2, B_3, B_4 and an identity matrix I (see, e.g., Bernstein (2005, p.247 and p.253)). Assumptions A.1(viii) and (ii) ensure the asymptotic positive definiteness of $\bar{A}$ and B. We can verify that $\left\|C'B^{-1}\right\|_{\text{sp}}^2 = \mu_{\max}\left(C'B^{-1}B^{-1}C\right) \leq \mu_{\max}\left(B^{-1}\right)\mu_{\max}\left(C'B^{-1}C\right) \leq \mu_{\max}\left(B^{-1}\right)\mu_{\max}(A)$ $= O_P(1)$ and $\|D\|_{\text{sp}} = O_P(1)$. By allowing $\|b\|$ and $\|\bar{\eta}\|$ to be sufficiently large, the linear terms $2b'C'B^{-1}\operatorname{vec}(D)$ and $-2\bar{\eta}'\operatorname{vec}(D)$ are dominated by the quadratic terms $b'\bar{A}b$ and $\bar{\eta}'B\bar{\eta}$, respectively. It follows that for any $\epsilon > 0$, there exists a large constant $\bar{L}$ such that

$$P\left\{\inf_{\|b\|=\bar{L},\|\bar{\eta}\|=\bar{L}} Q_\gamma\left(\beta^0 + a_{NT}b, \lambda^* + a_{NT}u\right) > Q_\gamma\left(\beta^0,\lambda^*\right)\right\} \geq 1-\epsilon$$

where $\bar{\eta} = \eta + B^{-1}Cb = \frac{1}{N^{1/2}}\operatorname{vec}(uH') + B^{-1}Cb$. Letting $\hat{b} = a_{NT}^{-1}(\hat{\beta} - \beta^0)$, $\hat{u} = a_{NT}^{-1}(\hat{\lambda} - \lambda^*)$, $\hat{\eta} = \frac{1}{N^{1/2}}\operatorname{vec}(\hat{u}H')$ and $\widehat{\bar{\eta}} = \hat{\eta} + B^{-1}C\hat{b}$, this further implies that either $||\hat{b}||$ or $||\widehat{\bar{\eta}}||$, or both must be stochastically bounded. We consider two cases: (a) $||\hat{b}||$ is stochastically bounded, and (b) $||\widehat{\bar{\eta}}||$ is stochastically bounded.

Suppose (a) holds. Then $\left\|\hat{\beta} - \beta^0\right\| = O_P(T^{-1/2})$ and the first part of the theorem follows. To prove the second part of the theorem, observe that

$$\begin{aligned}
0 &\geq a_{NT}^{-2}\left[Q_\gamma\left(\hat{\beta},\hat{\lambda}\right) - Q_\gamma\left(\beta^0,\lambda^*\right)\right] \geq a_{NT}^{-2}\Pi_{NT}\left(\hat{b},\hat{u}\right) + \text{s.m.} \\
&= \hat{b}'A\hat{b} + \operatorname{tr}\left(\frac{F^{0\prime}F^0}{T}\frac{H\hat{u}'\hat{u}H'}{N}\right) + 2\hat{b}'\frac{1}{NT}\sum_{t=1}^{T}X_{\cdot t}'\hat{u}H'F_t^0 - \frac{2}{NT^{1/2}}\operatorname{tr}\left(\boldsymbol{\varepsilon}F^0H\hat{u}'\right) + \text{s.m.} \\
&= \operatorname{tr}\left(\frac{F^{0\prime}F^0}{T}\frac{H\hat{u}'\hat{u}H'}{N}\right) + \frac{1}{\sqrt{N}}\|\hat{u}H'\|O_P(1) + O_P(1) + \text{s.m.},
\end{aligned}$$

where the last line follows from the fact that $\hat{b}'A\hat{b} \leq \mu_{\max}(A)||\hat{b}||^2 = O_P(1)$, $\left\|\hat{b}'\frac{1}{NT}\sum_{t=1}^{T}X_{\cdot t}'\hat{u}H'F_t^0\right\| \leq \frac{1}{\sqrt{N}}\|\hat{u}H'\|\frac{1}{N^{1/2}T}\sum_{t=1}^{T}\|X_{\cdot t}'\|\left\|F_t^0\right\| = \frac{1}{\sqrt{N}}\|\hat{u}H'\|O_P(1)$, and the analysis of $A_{2,2}(u)$ above. It follows that $\operatorname{tr}\left(\frac{F^{0\prime}F^0}{T}\frac{H\hat{u}'\hat{u}H'}{N}\right) = O_P(1)$, which further implies that $\operatorname{tr}(\frac{1}{N}H\hat{u}'\hat{u}H') = \frac{1}{N}||\hat{u}H'||^2 = O_P(1)$ as $\operatorname{tr}(B_1B_2) \geq \mu_{\min}(B_1)\operatorname{tr}(B_2)$ for any positive semidefinite matrix B_2 and symmetric matrix B_1 (see, e.g., Bernstein (2005, p.275)) and $T^{-1}F^{0\prime}F^0$ is asymptotically nonsingular by Assumption A.1(ii). Noting that $\hat{u} = \sqrt{T}(\hat{\lambda} - \lambda^0 H^{+\prime})$ and $HH^+ = I_{R_0}$, the last result implies that $\frac{1}{N}\left\|\hat{\lambda}H' - \lambda^0\right\|^2 = O_P\left(T^{-1}\right)$.

Now, suppose (b) holds. Note that

$$\begin{aligned} 0 &\geq a_{NT}^{-2}\left[Q_\gamma\left(\hat{\beta},\hat{\lambda}\right) - Q_\gamma\left(\beta^0,\lambda^*\right)\right] \geq \left[\hat{b}'\bar{A}\hat{b} + 2\hat{b}'C'B^{-1}\operatorname{vec}(D)\right] + \left[\widehat{\overline{\eta}}'B\widehat{\overline{\eta}} - 2\widehat{\overline{\eta}}'\operatorname{vec}(D)\right] + \text{s.m.} \\ &= \left[\hat{b}'\bar{A}\hat{b} - \hat{b}'C'B^{-1}\operatorname{vec}(D)\right] + O_P(1) + \text{s.m.}, \end{aligned}$$

which further implies that $||\hat{b}||$ is stochastically bounded by the positive definiteness of $\bar{A}$ and stochastic boundedness of $\|D\|_{\text{sp}}$ and $\left\|C'B^{-1}\right\|_{\text{sp}}$. Then $\|\eta\| = \left\|\widehat{\overline{\eta}} - B^{-1}C\hat{b}\right\| \leq \|\widehat{\overline{\eta}}\| + \left\|C'B^{-1}\right\|_{\text{sp}}\left\|\hat{b}\right\| = O_P(1)$. This, in conjunction with the fact that $HH^+ = I_{R_0}$, implies that

$$O_P(1) = \frac{1}{N}\|\hat{u}H'\|^2 = \frac{1}{N}\left\|a_{NT}^{-1}(\hat{\lambda} - \lambda^*)H'\right\|^2 = \frac{1}{N}\left\|a_{NT}^{-1}(\hat{\lambda}H' - \lambda^0)\right\|^2.$$

That is, $\frac{1}{N}\left\|\hat{\lambda}H' - \lambda^0\right\|^2 = O_P\left(T^{-1}\right)$. ■

B.2 Proof of Theorem 3.2

Let $\hat{\lambda}_{\cdot r}$ denotes the rth column of $\hat{\lambda}$ for $r = 1, ..., R$. We want to demonstrate that

$$P\left(\left|\hat{\beta}_k\right| = 0 \text{ and } \left\|\hat{\lambda}_{\cdot r}\right\| = 0 \text{ for } k = K_0+1,...,K, \text{ and } r = R_0+1,...,R\right) \to 1 \text{ as } (N,T) \to \infty.$$

Suppose that to the contrary, $\hat{\beta}_k \neq 0$ for some $k \in \{K_0+1, ..., K\}$ or $||\hat{\lambda}_{\cdot r}|| \neq 0$ for some $r \in \{R_0+1, ..., R\}$ for sufficiently large (N,T). Wlog assume that $\hat{\beta}_K \neq 0$ or $||\hat{\lambda}_{\cdot R}|| \neq 0$. If $\hat{\beta}_K \neq 0$, by the first order condition (FOC) with respect to β_K for our minimization problem, we have

$$\begin{aligned} 0 &= -\frac{2}{NT^{1/2}}\operatorname{tr}\left[\mathbf{X}_K'\left(\mathbf{Y} - \sum_{k=1}^K\hat{\beta}_k\mathbf{X}_k - \hat{\lambda}\hat{F}'\right)\right] + \frac{T^{1/2}\gamma_{1NT}}{\left|\tilde{\beta}_K^c\right|^{\kappa_1}}\frac{\hat{\beta}_K}{\left|\hat{\beta}_K\right|} \qquad \text{(B.2)} \\ &= -\frac{2}{NT^{1/2}}\operatorname{tr}\left(\mathbf{X}_K'\boldsymbol{\varepsilon}\right) + \frac{2}{NT}\sum_{k=1}^K\operatorname{tr}\left(\mathbf{X}_K'\mathbf{X}_k\right)\sqrt{T}\left(\hat{\beta}_k - \beta_k^0\right) \\ &\quad + \frac{2}{NT^{1/2}}\operatorname{tr}\left[\mathbf{X}_K'\left(\hat{\lambda}\hat{F}' - \lambda^0F^{0\prime}\right)\right] + \frac{T^{1/2}\gamma_{1NT}}{\left|\tilde{\beta}_K^c\right|^{\kappa_1}}\frac{\hat{\beta}_K}{\left|\hat{\beta}_K\right|} \\ &\equiv -2B_1 + 2B_2 + 2B_3 + 2B_4, \text{ say.} \end{aligned}$$

Note that $B_1 = O_P(N^{-1/2})$ by Assumption A.1(vi). For B_2, using $|\operatorname{tr}(C_1C_2)| \leq \|C_1\|\|C_2\|$ and $b'Bb \leq \mu_{\max}(B)b'b$ for any conformable matrices C_1, C_2, b and B, we have by Assumptions A.1(iv) and (viii) and Theorem 3.1,

$$\begin{aligned} |B_2|^2 &\leq \frac{1}{N^2T}\|\mathbf{X}_K\|^2\operatorname{tr}\left(\sum_{k=1}^K\sum_{l=1}^K\left(\hat{\beta}_k - \beta_k^0\right)\mathbf{X}_k'\mathbf{X}_l\left(\hat{\beta}_l - \beta_l^0\right)\right) \\ &= \frac{1}{N}\|\mathbf{X}_K\|^2\left(\hat{\beta} - \beta^0\right)'\bar{W}_{NT}\left(\hat{\beta} - \beta^0\right) \\ &\leq \left(\frac{1}{NT}\|\mathbf{X}_K\|^2\right)\mu_{\max}\left(\bar{W}_{NT}\right)\left(T\left\|\hat{\beta} - \beta^0\right\|^2\right) = O_P(1)\,O_P(1)\,O_P(1) = O_P(1), \end{aligned}$$

where we recall that $\bar{W}_{NT} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}X_{it}'$. In addition, $|B_3| \leq \frac{1}{NT^{1/2}}|\text{tr}[\mathbf{X}_K'(\hat{\lambda}H' - \lambda^0)F^{0\prime}]| + \frac{1}{NT^{1/2}}|\text{tr}[\mathbf{X}_K'\hat{\lambda}(\hat{F} - F^0H)']| \equiv B_{3,1} + B_{3,2}$. By Assumptions A.1(ii) and (iv) and Theorem 3.1, $B_{3,1} \leq (NT)^{-1/2}\|\mathbf{X}_K\|\{T^{1/2}N^{-1/2}||\hat{\lambda}H' - \lambda^0||\}T^{-1/2}\|F^0\| = O_P(1)$. Similarly, $B_{3,2} = O_P(1)$ by Assumption A.1(iv), Lemma A.2(i) and the remark after Theorem 3.1. It follows that $B_3 = O_P(1)$ and $-B_1 + B_2 + B_3 = O_P(1)$. Noting that $\tilde{\beta}_K^c = O_P((NT)^{-1/2})$ by Assumption A.1(i), $|B_4| = \frac{T^{1/2}\gamma_{1NT}}{|\tilde{\beta}_K^c|^{\kappa_1}}$ is explosive in probability because $(NT)^{\kappa_1/2}T^{1/2}\gamma_{1NT} \to \infty$ by Assumption A.3(iii). This implies that (B.2) cannot be true for sufficiently large N and T. Consequently, w.p.a.1 $\hat{\beta}_K$ must be in a position where $|\beta_K|$ is not differentiable, i.e., $P(|\hat{\beta}_K| = 0) \to 1$ as $(N,T) \to \infty$.

Let $\lambda_{i,r}$ and $\hat{\lambda}_{i,r}$ denote the rth elements of λ_i and $\hat{\lambda}_i$, respectively. If $||\hat{\lambda}_{\cdot R}|| \neq 0$, the FOC for $\lambda_{i,R}$ implies that for all $i = 1,...,N$, we have

$$\begin{aligned}
0 &= \frac{-2}{T^{1/2}}\sum_{t=1}^{T}[\hat{F}_{t,R}(Y_{it} - X_{it}'\hat{\beta}_k - \hat{\lambda}_i'\hat{F}_t)] + \frac{\sqrt{NT}\gamma_{2NT}}{|\tau_R|^{\kappa_2}}\frac{\hat{\lambda}_{i,R}}{||\hat{\lambda}_{\cdot R}||} \qquad \text{(B.3)}\\
&= -\frac{2}{T^{1/2}}\sum_{t=1}^{T}\hat{F}_{t,R}\varepsilon_{it} + \frac{2}{T}\sum_{t=1}^{T}\hat{F}_{t,R}X_{it}'\sqrt{T}(\hat{\beta} - \beta^0)\\
&\quad + \frac{2}{T^{1/2}}\sum_{t=1}^{T}\hat{F}_{t,R}(\hat{\lambda}_i'\hat{F}_t - \lambda_i^{0\prime}H^{+\prime}H'F_t^0) + \frac{\sqrt{NT}\gamma_{2NT}}{|\tau_R|^{\kappa_2}}\frac{\hat{\lambda}_{i,R}}{||\hat{\lambda}_{\cdot R}||}\\
&\equiv 2d_{1i} + 2d_{2i} + 2d_{3i} + d_{4i}, \text{ say,}
\end{aligned}$$

where $\hat{F}_{t,R}$ denotes the Rth element of $\hat{F}_t$. Observe that $N^{-1}\sum_{i=1}^{N}(d_{1i} + d_{2i} + d_{3i})^2 \leq 3N^{-1}\sum_{i=1}^{N}(d_{1i}^2 + d_{2i}^2 + d_{3i}^2)$. By the C_r inequality, Lemma A.2(i), Assumption A.2(i), we have

$$\begin{aligned}
N^{-1}\sum_{i=1}^{N}d_{1i}^2 &\leq 2N^{-1}T^{-1}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}\left(\hat{F}_t - H'F_t^0\right)_R\varepsilon_{it}\right]^2 + 2N^{-1}T^{-1}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}\left(H'F_t^0\right)_R\varepsilon_{it}\right]^2\\
&\leq 2N^{-1}T^{-1}\sum_{i=1}^{N}\left\|\varepsilon_i'(\hat{F}_{\cdot R} - F^0H_{\cdot R})\right\|^2 + 2N^{-1}T^{-1}\sum_{i=1}^{N}\left\|\varepsilon_i'F^0\right\|^2\|H_{\cdot R}\| \equiv D_{1,1} + D_{1,2},
\end{aligned}$$

where $(b)_R$ and $B_{\cdot R}$ denote the Rth element of the $R \times 1$ vector b and the Rth column of matrix B, respectively. $D_{1,2} = O_P\left(\delta_{NT}^{-1}\right)$ by the fact that $N^{-1}T^{-1}\sum_{i=1}^{N}\left\|\varepsilon_i'F^0\right\|^2 = O_P(1)$ under Assumptions A.1(vii) and (iv) and $\|H_{\cdot R}\| = O_P\left(\delta_{NT}^{-1}\right)$ by Lemma A.2(ii). Using the decomposition for $\hat{F} - F^0H$ in the proof of Lemma A.2(i), we can readily show that $D_{1,1} = O_P(1)$. It follows that $N^{-1}\sum_{i=1}^{N}d_{1i}^2 = O_P(1)$. By CS inequality, Theorem 3.1, Lemma A.2(i), and Assumption A.1(viii),

$$\begin{aligned}
N^{-1}\sum_{i=1}^{N}d_{2i}^2 &\leq N^{-1}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_{t,R}\right)^2\ \frac{1}{T}\sum_{t=1}^{T}\left[X_{it}'\sqrt{T}(\hat{\beta} - \beta^0)\right]^2\\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_{t,R}\right)^2\sqrt{T}(\hat{\beta} - \beta^0)'\bar{W}_{NT}\sqrt{T}(\hat{\beta} - \beta^0)\\
&\leq T||\hat{\beta} - \beta^0||^2\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{F}_{t,R}\right\|^2\mu_{\max}\left(\bar{W}_{NT}\right) = O_P(1).
\end{aligned}$$

Noting that $\sum_{t=1}^{T} \hat{F}_{t,R}(\hat{\lambda}_i' \hat{F}_t - \lambda_i^{0\prime} H^{+\prime} H' F_t^0) = \hat{F}_{\cdot R}' \hat{F}(\hat{\lambda}_i - H^+ \lambda_i^0) + \hat{F}_{\cdot R}'(\hat{F} - F^0 H) H^+ \lambda_i^0$, we have

$$\begin{aligned} N^{-1} \sum_{i=1}^{N} d_{3i}^2 &\leq 2N^{-1}T^{-1} \sum_{i=1}^{N} \left[\hat{F}_{\cdot R}' \hat{F} \left(\hat{\lambda}_i - H^+ \lambda_i^0 \right) \right]^2 + 2N^{-1}T^{-1} \sum_{i=1}^{N} \left[\hat{F}_{\cdot R}' \left(\hat{F} - F^0 H \right) H^+ \lambda_i^0 \right]^2 \\ &\equiv 2D_{3,1} + 2D_{3,2}, \text{ say.} \end{aligned}$$

Noting that $\hat{F}_{\cdot R}' \hat{F} = (\hat{F}_{\cdot R} - F^0 H_{\cdot R})'(\hat{F} - F^0 H) + (\hat{F}_{\cdot R} - F^0 H_{\cdot R})' F^0 H + H_{\cdot R}' F^{0\prime}(\hat{F} - F^0 H) + H_{\cdot R}' F^{0\prime} F^0 H$, we have by the C_r inequality,

$$\begin{aligned} D_{3,1} &\leq 4N^{-1}T^{-1} \sum_{i=1}^{N} \left[(\hat{F}_{\cdot R} - F^0 H_{\cdot R})'(\hat{F} - F^0 H) \left(\hat{\lambda}_i - H^+ \lambda_i^0 \right) \right]^2 \\ &\quad + 4N^{-1}T^{-1} \sum_{i=1}^{N} \left[(\hat{F}_{\cdot R} - F^0 H_{\cdot R})' F^0 H \left(\hat{\lambda}_i - H^+ \lambda_i^0 \right) \right]^2 \\ &\quad + 4N^{-1}T^{-1} \sum_{i=1}^{N} \left[H_{\cdot R}' F^{0\prime}(\hat{F} - F^0 H) \left(\hat{\lambda}_i - H^+ \lambda_i^0 \right) \right]^2 + 4N^{-1}T^{-1} \sum_{i=1}^{N} \left[H_{\cdot R}' F^{0\prime} F^0 H \left(\hat{\lambda}_i - H^+ \lambda_i^0 \right) \right]^2 \\ &\equiv 4D_{3,1,1} + 4D_{3,1,2} + 4D_{3,1,3} + 4D_{3,1,4}, \text{ say.} \end{aligned}$$

By Lemma A.2(i) and the remark after Theorem 3.1, $D_{3,1,1} \leq T\{T^{-1}||\hat{F}_{\cdot R} - F^0 H_{\cdot R}||^2\}\{T^{-1}||\hat{F} - F^0 H||^2\}$ $\{N^{-1}||\hat{\lambda} - \lambda^0 H^{+\prime}||^2\} = TO_P\left(\delta_{NT}^{-2}\right) O_P\left(\delta_{NT}^{-2}\right) O_P(1) = o_P(1)$ under Assumption A.3(i). Similarly, by Lemma A.2, Theorem 3.1 and the remark after it, we have

$$\begin{aligned} D_{3,1,2} &\leq T^{-1} \left\| (\hat{F}_{\cdot R} - F^0 H_{\cdot R})' F^0 \right\|^2 N^{-1} \sum_{i=1}^{N} \left(H \hat{\lambda}_i - \lambda_i^0 \right)^2 = O_P\left(T\delta_{NT}^{-2} \right) O_P\left(T^{-1} \right), \\ D_{3,1,3} &\leq \left\| H_{\cdot R} \right\|^2 T^{-1} \left\| F^{0\prime}(\hat{F} - F^0 H) \right\|^2 N^{-1} \sum_{i=1}^{N} \left\| \hat{\lambda}_i - H^+ \lambda_i^0 \right\|^2 = O_P\left(\delta_{NT}^{-2} \right) O_P\left(T\delta_{NT}^{-2} \right) O_P(1), \\ D_{3,1,4} &\leq \left\| H_{\cdot R} \right\|^2 T^{-1} \left\| F^{0\prime} F^0 \right\|^2 N^{-1} \sum_{i=1}^{N} \left\| H \hat{\lambda}_i - \lambda_i^0 \right\|^2 = O_P\left(\delta_{NT}^{-2} \right) O_P(T) O_P\left(T^{-1} \right). \end{aligned}$$

It follows that $D_{3,1} = o_P(1)$. In addition, by the triangle inequality and Lemma A.2(iv),

$$D_{3,2} \leq T^{-1}||\hat{F}_{\cdot R}'(\hat{F} - F^0 H)||^2 N^{-1} \sum_{i=1}^{N} \left\| \lambda_i^0 \right\|^2 \left\| H^+ \right\|^2 = O_P\left(T\delta_{NT}^{-4} \right) O_P(1) = o_P(1).$$

It follows that $N^{-1} \sum_{i=1}^{N} d_{3i}^2 = o_P(1)$. Consequently we have shown that

$$N^{-1} \sum_{i=1}^{N} (d_{1i} + d_{2i} + d_{3i})^2 = O_P(1). \tag{B.4}$$

Now, we study $N^{-1} \sum_{i=1}^{N} d_{4i}^2$. In view of the fact that $\tau_R = O_P(N^{-1/2})$ by Lemma A.2(vii), $N^{-1} \sum_{i=1}^{N} d_{4i}^2$ $= \left(\frac{\sqrt{T}\gamma_{2NT}}{|\tau_R|^{\kappa_2}} \right)^2$ is explosive in probability because $N^{\kappa_2/2} T^{1/2} \gamma_{2NT} \to \infty$ by Assumption A.3(iii). This, in conjunction with (B.4), implies that $N^{-1} \sum_{i=1}^{N} d_{4i}^2 \gg 4N^{-1} \sum_{i=1}^{N} (d_{1i} + d_{2i} + d_{3i})^2$ so that (B.3) cannot be true for all i for sufficiently large N and T. Then we conclude that w.p.a.1, $||\hat{\lambda}_{\cdot R}||$ must be in a position where $\|\lambda_{\cdot R}\|$ is not differentiable. Consequently $P(||\hat{\lambda}_{\cdot R}|| = 0) \to 1$ as $(N, T) \to \infty$. ■

B.3 Proof of Theorem 3.3

Define

$$D(\tilde{\beta}^c_{(1)};\hat{\beta}_{(1)}) \equiv \left(\frac{\gamma_{1NT}}{2\left|\tilde{\beta}^c_1\right|^{\kappa_1}}\frac{\hat{\beta}_1}{\left|\hat{\beta}_1\right|},...,\frac{\gamma_{1NT}}{2\left|\tilde{\beta}^c_{K_0}\right|^{\kappa_1}}\frac{\hat{\beta}_{K_0}}{\left|\hat{\beta}_{K_0}\right|}\right)',$$

$$D_i(\tilde{\lambda}_{(1)};\hat{\lambda}_{(1)}) \equiv \left(\frac{\sqrt{N}\gamma_{2NT}}{2\tau_1^{\kappa_2}}\frac{\hat{\lambda}_{i,1}}{\left\|\hat{\lambda}_{\cdot 1}\right\|},...\frac{\sqrt{N}\gamma_{2NT}}{2\tau_{R_0}^{\kappa_2}}\frac{\hat{\lambda}_{i,R_0}}{\left\|\hat{\lambda}_{\cdot R_0}\right\|}\right)' \text{ for } i=1,...,N.$$

Let $D(\tilde{\beta}^c_{(1)}) \equiv D(\tilde{\beta}^c_{(1)};\tilde{\beta}^c_{(1)})$ and $\tilde{D}_i(\tilde{\lambda}_{(1)}) = D_i(\tilde{\lambda}_{(1)};\hat{\lambda}_{(1)})$, where, e.g., $\hat{\lambda}_{(1)} = (\hat{\lambda}_{\cdot 1},...,\hat{\lambda}_{\cdot R_0})$. W.p.a.1, $\hat{\beta}_k/|\hat{\beta}_k| = \tilde{\beta}^c_k/|\tilde{\beta}^c_k|$ for $k=1,....,K_0$ by the consistency of $\hat{\beta}_k$ and $\tilde{\beta}^c_k$ and the fact that $\beta^0_k \neq 0$ for $k=1,....,K_0$. So $D(\tilde{\beta}^c_{(1)};\hat{\beta}_{(1)}) = D(\tilde{\beta}^c_{(1)})$ w.p.a.1. It follows from Theorem 3.2 and the FOCs, w.p.a.1 we have

$$\begin{cases} 0 = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T X_{it(1)}\left[Y_{it} - X'_{it(1)}\hat{\beta}_{(1)} - \hat{F}'_{t(1)}\hat{\lambda}_{i(1)}\right] - D(\tilde{\beta}^c_{(1)}), \\ 0 = \frac{1}{T}\sum_{t=1}^T \hat{F}_{t(1)}\left[Y_{it} - X'_{it(1)}\hat{\beta}_{(1)} - \hat{F}'_{t(1)}\hat{\lambda}_{i(1)}\right] - \tilde{D}_i(\tilde{\lambda}_{(1)}) \text{ for } i=1,...,N. \end{cases}$$

Solving for $\hat{\beta}_{(1)}$ and $\hat{\lambda}_{i(1)}$ yields

$$\begin{cases} \hat{\beta}_{(1)} = D^{-1}_{\hat{F}_{(1)}}\left[\frac{1}{NT}\sum_{i=1}^N X'_{i(1)}M_{\hat{F}_{(1)}}Y_i - D(\tilde{\beta}^c_{(1)}) + \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T X_{it(1)}\hat{F}'_{t(1)}\hat{\Sigma}^{-1}_{\hat{F}_{(1)}}\tilde{D}_i(\tilde{\lambda}_{(1)})\right], \\ \hat{\lambda}_{i(1)} = \hat{\Sigma}^{-1}_{\hat{F}_{(1)}}\left[\frac{1}{T}\sum_{t=1}^T \hat{F}_{t(1)}(Y_{it} - X'_{it(1)}\hat{\beta}^c_{(1)}) - \tilde{D}_i(\tilde{\lambda}_{(1)})\right] \text{ for } i=1,2,...,N, \end{cases}$$

where $\hat{D}_{\hat{F}_{(1)}} = \frac{1}{NT}\sum_{i=1}^N X'_{i(1)}M_{\hat{F}_{(1)}}X_{i(1)}$ and $\hat{\Sigma}_{\hat{F}_{(1)}} = \frac{1}{T}\sum_{t=1}^T \hat{F}_{t(1)}\hat{F}'_{t(1)}$. By Lemma A.2(vi) and Assumption A.3(i), $||\hat{D}_{\hat{F}_{(1)}} - \hat{D}_{F^*_{(1)}}||_{\text{sp}} \leq ||\hat{D}_{\hat{F}_{(1)}} - \hat{D}_{F^*_{(1)}}|| = \frac{1}{NT}\sum_{i=1}^N ||X_{i(1)}||^2||P_{\hat{F}_{(1)}} - P_{F^*_{(1)}}|| = O_P(K_0\delta^{-1}_{NT}) = o_P(1)$. In view of the fact that $M_{F^*_{(1)}} = M_{F^0}$, we have $||\hat{D}_{F^*_{(1)}} - D_{F^0}||_{\text{sp}} = ||\hat{D}_{F^0} - D_{F^0}||_{\text{sp}} + o_P(1)$ by Assumption A.4(i). Then $||\hat{D}_{\hat{F}_{(1)}} - D_{F^0}||_{\text{sp}} = o_P(1)$ by the triangle inequality and $||D^{-1}_{\hat{F}_{(1)}} - D^{-1}_{F^0}||_{\text{sp}} = o_P(1)$. By Lemmas A.2(v) and A.1(iv) and Assumption A.1(ii)-(iii), $\hat{\Sigma}^{-1}_{\hat{F}_{(1)}} = O_P(1)$. By Assumption A.6(ii) and the fact that $||\hat{\lambda}_{\cdot r}||^{-1} = O_P(N^{-1/2})$ for $r=1,...,R_0$, $\left\|D(\tilde{\beta}^c_{(1)})\right\| = O_P(K_0^{1/2}\gamma_{1NT}) = o_P((NT)^{-1/2})$ and $\left\|\tilde{D}_i(\tilde{\lambda}_{(1)})\right\| = O_P(\gamma_{2NT}) = o_P((NT)^{-1/2})$. With these results, we can readily show that

$$\begin{cases} \hat{\beta}_{(1)} = D^{-1}_{\hat{F}_{(1)}}\frac{1}{NT}\sum_{i=1}^N X'_{i(1)}M_{\hat{F}_{(1)}}Y_i + \mathbf{o}_P\left((NT)^{-1/2}\right), \\ \hat{\lambda}_{i(1)} = \hat{\Sigma}^{-1}_{\hat{F}_{(1)}}\frac{1}{T}\sum_{t=1}^T \hat{F}_{t(1)}\left(Y_{it} - X'_{it(1)}\hat{\beta}_{(1)}\right) + \mathbf{o}_P\left((NT)^{-1/2}\right) \text{ for } i=1,2,...,N. \end{cases}$$

Below we study the asymptotic distributions of $\hat{\beta}_{(1)}$ and $\hat{\lambda}_{i(1)}$ in turn.

Asymptotic distribution of $\hat{\beta}_{(1)}$. Noting that $Y_i = \hat{F}_{(1)}\lambda^*_{i(1)} + X_{i(1)}\beta^0_{(1)} + \varepsilon_i + (F^*_{(1)} - \hat{F}_{(1)})\lambda^*_{i(1)}$, we have

$$\begin{aligned} \sqrt{NT}\mathbb{C}_K\left(\hat{\beta}_{(1)} - \beta^0_{(1)}\right) &= \mathbb{C}_K D^{-1}_{\hat{F}_{(1)}}\frac{1}{\sqrt{NT}}\sum_{i=1}^N X'_{i(1)}M_{F^0}\varepsilon_i + \mathbb{C}_K D^{-1}_{\hat{F}_{(1)}}\frac{1}{\sqrt{NT}}\sum_{i=1}^N X'_{i(1)}M_{\hat{F}_{(1)}}(F^*_{(1)} - \hat{F}_{(1)})\lambda^*_{i(1)} \\ &\quad + \mathbb{C}_K D^{-1}_{\hat{F}_{(1)}}\frac{1}{\sqrt{NT}}\sum_{i=1}^N X'_{i(1)}\left(M_{\hat{F}_{(1)}} - M_{F^*_{(1)}}\right)\varepsilon_i + o_P(1) \\ &\equiv S_{1NT} + S_{2NT} + S_{3NT} + o_P(1), \text{ say.} \end{aligned}$$

By Propositions B.1-B.2 below and the fact that $||D_{\hat{F}_{(1)}}^{-1} - D_{F^0}^{-1}||_{\mathrm{sp}} = o_P(1)$, we have $S_{2NT} = \mathbb{C}_K D_{\hat{F}_{(1)}}^{-1}(\mathbb{B}_{1NT} - \mathbb{B}_{2NT} - \mathbb{V}_{1NT} + \mathbb{V}_{2NT}) + o_P(1)$, and $S_{3NT} = -\mathbb{C}_K D_{\hat{F}_{(1)}}^{-1}\mathbb{B}_{3NT} + o_P(1)$, where $\|\mathbb{B}_{1NT}\| = O_P(\sqrt{KT/N})$, $\|\mathbb{B}_{2NT}\| = O_P(\sqrt{K} + \sqrt{KN/T})$, $\|\mathbb{B}_{3NT}\| = O_P(\sqrt{KT/N})$, and $\|\mathbb{V}_{lNT}\| = O_P\left(\sqrt{K}\right)$ for $l = 1, 2$. Consequently,

$$
\begin{aligned}
&\sqrt{NT}\mathbb{C}_K\left(\hat{\beta}_{(1)} - \beta_{(1)}^0\right) - \mathbb{C}_K D_{\hat{F}_{(1)}}^{-1}(\mathbb{B}_{1NT} - \mathbb{B}_{2NT} - \mathbb{B}_{3NT}) \\
= \; & \mathbb{C}_K D_{\hat{F}_{(1)}}^{-1}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} X_{i(1)}' M_{F^0}\varepsilon_i - \mathbb{V}_{1NT} + \mathbb{V}_{2NT}\right) + o_P(1).
\end{aligned}
$$

Following Moon and Weidner (2014a, 2014b) and as demonstrated in the supplementary Appendix F , we have $\sqrt{NT}(\tilde{\beta}^c - \beta^0) = W_{NT}^{-1}V_{NT} + \mathbf{o}_P(1)$, where $W_{NT} = \frac{1}{NT}\sum_{i=1}^N \tilde{X}_i' M_{F^0}\tilde{X}_i$, $\tilde{X}_i = X_i - \mathcal{X}_{i2NT}$, $V_{NT} = \frac{1}{\sqrt{NT}}\sum_{i=1}^N [X_i - P_{F^0}E_{\mathcal{D}}(X_i) - M_{F^0}\mathcal{X}_{i2NT}]'\varepsilon_i$, and $\mathcal{X}_{i2NT} \equiv \frac{1}{N}\sum_{j=1}^N \lambda_i^{0\prime}\left(N^{-1}\lambda^{0\prime}\lambda^0\right)^{-1}\lambda_j^0 X_j$.[18] It follows that $\frac{1}{\sqrt{NT}}\sum_{i=1}^N X_{i(1)}' M_{F^0}\varepsilon_i - \mathbb{V}_{1NT} + \mathbb{V}_{2NT} = J_{NT} + \mathbf{o}_P(1)$, where

$$
J_{NT} = \frac{1}{\sqrt{NT}}\sum_{i=1}^N \left\{M_{F^0}X_{i(1)} - M_{F^0}\mathcal{X}_{i1NT} + [X_i - P_{F^0}E_{\mathcal{D}}(X_i) - M_{F^0}\mathcal{X}_{i2NT}]\,W_{NT}^{-1}C_{NT}\right\}'\varepsilon_i
$$

and $\mathcal{X}_{i1NT}$ is defined in Remark 4. In the presence of lagged dependent variables, J_{NT} does not center around 0 asymptotically and it contributes to both the asymptotic bias and variance of $\hat{\beta}_{(1)}$. We make the following decomposition: $J_{NT} = \frac{1}{\sqrt{NT}}\sum_{i=1}^N Z_i'\varepsilon_i - \mathbb{B}_{4NT}$, where $Z_i \equiv X_{i(1)} - P_{F^0}E_{\mathcal{D}}(X_{i(1)}) - M_{F^0}\mathcal{X}_{i1NT} + [X_i - P_{F^0}E_{\mathcal{D}}(X_i) - M_{F^0}\mathcal{X}_{i2NT}]\;W_{NT}^{-1}C_{NT}$ and $\mathbb{B}_{4NT} \equiv \frac{1}{\sqrt{NT}}\sum_{i=1}^N [X_{i(1)} - E_{\mathcal{D}}(X_{i(1)})]' P_{F^0}\varepsilon_i$. Then by Assumptions A.4(i)-(ii) and A.5(i)

$$
\begin{aligned}
\mathbb{C}_{K_0}\left[\sqrt{NT}\left(\hat{\beta}_{(1)} - \beta_{(1)}^0\right) - \mathbb{B}_{NT}\right] &= \frac{1}{\sqrt{NT}}\mathbb{C}_{K_0} D_{\hat{F}_{(1)}}^{-1}\sum_{i=1}^N Z_i'\varepsilon_i + o_P(1) \\
&= \frac{1}{\sqrt{NT}}\mathbb{C}_{K_0} D_{F^0}^{-1}\sum_{i=1}^N Z_i'\varepsilon_i + o_P(1) \xrightarrow{d} N\left(0, \lim_{(N,T)\to\infty}\mathbb{C}_{K_0}\mathbb{V}_{NT}\mathbb{C}_{K_0}'\right),
\end{aligned}
$$

where $\mathbb{B}_{NT} = D_{\hat{F}_{(1)}}^{-1}(\mathbb{B}_{1NT} - \mathbb{B}_{2NT} - \mathbb{B}_{3NT} - \mathbb{B}_{4NT})$ and $\mathbb{V}_{NT} = D_{F^0}^{-1}\Theta_{NT}D_{F^0}^{-1}$.

Asymptotic distribution of $\hat{\lambda}_{i(1)}$. Noting that $Y_i - X_{it(1)}'\hat{\beta}_{(1)} = \hat{F}_{(1)}\lambda_{i(1)}^* + \varepsilon_i - X_{i(1)}(\hat{\beta}_{(1)} - \beta_{(1)}^0) +$

[18] Let $C_{NT}^{(1)} = \frac{1}{\sqrt{NT}}\sum_{i=1}^N \tilde{X}_i' M_{F^0}\varepsilon_i$. Its kth element corresponds to $\frac{1}{\sqrt{NT}}C^{(1)}\left(\lambda^0, f^0, X_k, e\right)$ defined in Moon and Weidner (2014b, Section 4.1) which contributes to both the asymptotic bias and variance. We make the following decomposition:

$$
\begin{aligned}
C_{NT}^{(1)} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^N (X_i - M_{F^0}\mathcal{X}_{i2NT})'\varepsilon_i - \frac{1}{\sqrt{NT}}\sum_{i=1}^N X_i' P_{F^0}\varepsilon_i \\
&= \frac{1}{\sqrt{NT}}\sum_{i=1}^N [X_i - P_{F^0}E_{\mathcal{D}}(X_i) - M_{F^0}\mathcal{X}_{i2NT}]'\varepsilon_i - \frac{1}{\sqrt{NT}}\sum_{i=1}^N [X_i - E_{\mathcal{D}}(X_i)]' P_{F^0}\varepsilon_i
\end{aligned}
$$

where the first term is V_{NT} and contributes to the asymptotic variance and the second term can be corrected as in Moon and Weidner (2014b); see also the proof of Corollary 3.4 for the correction of $\mathbb{B}_{4NT}$. In the absence of lagged dependent variables, one can replace V_{NT} by $C_{NT}^{(1)}$ as in Bai (2009).

$(F_{(1)}^{*}-\hat{F}_{(1)})\lambda_{i(1)}^{*}$, we have

$$\begin{aligned}\sqrt{T}\left(\hat{\lambda}_{i(1)}-\lambda_{i(1)}^{*}\right) &= \hat{\Sigma}_{\hat{F}_{(1)}}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{F}_{t(1)}\varepsilon_{it}-\hat{\Sigma}_{\hat{F}_{(1)}}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{F}_{t(1)}X_{it(1)}^{\prime}\left(\hat{\beta}_{(1)}-\beta_{(1)}^{0}\right)\\ &\quad+\hat{\Sigma}_{\hat{F}_{(1)}}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{F}_{t(1)}\left(F_{t(1)}^{*}-\hat{F}_{t(1)}\right)^{\prime}\lambda_{i(1)}^{*}+o_{P}\left(N^{-1/2}\right)\\ &= e_{1i}-e_{2i}+e_{3i}+o_{P}\left(N^{-1/2}\right),\text{ say.}\end{aligned}$$

By Lemmas A.2(v) and A.1(iv), $\hat{\Sigma}_{\hat{F}_{(1)}}=\frac{1}{T}H_{(1)}^{\prime}F^{0\prime}F^{0}H_{(1)}+O_{P}\left(\delta_{NT}^{-2}\right)=\Sigma_{F_{(1)}^{*}}+o_{P}\left(1\right).$ By Lemma A.2(i) and Assumptions A.1(ii) and (iv),

$$\begin{aligned}\left\Vert\frac{1}{T}\sum_{t=1}^{T}\hat{F}_{t(1)}X_{it(1)}^{\prime}\right\Vert &\leq \left\Vert H_{(1)}^{\prime}\frac{1}{T}\sum_{t=1}^{T}F_{t}^{0}X_{it(1)}^{\prime}\right\Vert+\left\Vert\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_{t(1)}-H_{(1)}^{\prime}F_{t}^{0}\right)X_{it(1)}^{\prime}\right\Vert\\ &= O_{P}\left(K_{0}^{1/2}\right)+\left\{\frac{1}{T}\sum_{t=1}^{T}\left\Vert\hat{F}_{t(1)}-H_{(1)}^{\prime}F_{t}^{0}\right\Vert^{2}\frac{1}{T}\sum_{t=1}^{T}\left\Vert X_{it(1)}\right\Vert^{2}\right\}^{1/2}\\ &= O_{P}\left(K_{0}^{1/2}\right)+O_{P}\left(K_{0}^{1/2}\delta_{NT}^{-1}\right)=O_{P}\left(K_{0}^{1/2}\right).\end{aligned}$$

By the study of the asymptotic distribution of $\hat{\beta}_{(1)}$, $\left\Vert\hat{\beta}_{(1)}-\beta_{(1)}^{0}\right\Vert=O_{P}(K_{0}^{1/2}\delta_{NT}^{-2})$. It follows that $e_{2i}=O_{P}\left(K_{0}T^{1/2}\delta_{NT}^{-2}\right)=o_{P}\left(1\right).$ By Lemma A.2(iv) and Assumption A.3(i)

$$\begin{aligned}\left\Vert e_{3i}\right\Vert &\leq \sqrt{T}\left\Vert\hat{\Sigma}_{\hat{F}_{(1)}}^{-1}\right\Vert\frac{1}{T}\left\Vert\hat{F}_{(1)}^{\prime}(F^{0}H_{(1)}-\hat{F}_{(1)})\right\Vert\left\Vert\lambda_{i(1)}^{0}\right\Vert\\ &= \sqrt{T}O_{P}\left(1\right)O_{P}(\delta_{NT}^{-2}+K\left(NT\right)^{-1/2})O_{P}\left(1\right)=o_{P}\left(1\right).\end{aligned}$$

In addition, by Lemma A.3(iv) and Assumptions A.3(i) and A.5(ii), $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{F}_{t(1)}\varepsilon_{it}=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_{t(1)}^{*}\varepsilon_{it}+o_{P}\left(1\right)\xrightarrow{d}N(0,\Theta_{i,F_{(1)}^{*}})$. It follows that $\sqrt{T}(\hat{\lambda}_{i(1)}-\lambda_{i(1)}^{*})\xrightarrow{d}N(0,\Sigma_{F_{(1)}^{*}}^{-1}\Theta_{i,F_{(1)}^{*}}\Sigma_{F_{(1)}^{*}}^{-1})$ by Slutsky theorem. ■

Proposition B.1 *Suppose that the conditions of Theorem 3.3 hold. Then* $s_{NT}\equiv\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}X_{i(1)}^{\prime}M_{\hat{F}_{(1)}}(F_{(1)}^{*}-\hat{F}_{(1)})\lambda_{i(1)}^{*}=\mathbb{B}_{1NT}-\mathbb{B}_{2NT}-\mathbb{V}_{1NT}+\mathbb{V}_{2NT}+\mathbf{o}_{P}\left(1\right),$ *where* $\mathbb{B}_{1NT}$, $\mathbb{B}_{2NT}$, $\mathbb{V}_{1NT}$, *and* $\mathbb{V}_{2NT}$ *are defined in Section 3.3.*

Proof. Note that

$$\begin{aligned}s_{NT} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}X_{i(1)}^{\prime}(M_{\hat{F}_{(1)}}-M_{F_{(1)}^{*}})(F_{(1)}^{*}-\hat{F}_{(1)})\lambda_{i(1)}^{*}+\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}X_{i(1)}^{\prime}M_{F_{(1)}^{*}}(F_{(1)}^{*}-\hat{F}_{(1)})\lambda_{i(1)}^{*}\\ &\equiv s_{1NT}+s_{2NT},\text{ say.}\end{aligned}$$

By Lemma A.3(i), $s_{1NT}=\mathbb{B}_{1NT}+\mathbf{o}_{P}\left(1\right).$ By (C.2) in the supplementary appendix and the fact that $M_{F^{0}}=M_{F_{(1)}^{*}}$, $s_{2NT}=-\sum_{l=1}^{8}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}X_{i(1)}^{\prime}M_{F^{0}}a_{l(1)}\lambda_{i(1)}^{*}\equiv-\sum_{l=1}^{8}s_{2NT,l}$, say. We dispense with the terms that are easy to analyze first. First, $s_{2NT,2}=s_{2NT,5}=0$ as $M_{F^{0}}F^{0}=0$. Next, we want to show that $s_{2NT,l}=o_{P}\left(1\right)$ for $l=4$, 7, and 8. Let $c_{K_{0}}=(c_{1K_{0}},...,c_{K_{0}K_{0}})^{\prime}$ be an arbitrary $K_{0}\times1$ nonrandom vector with $\left\Vert c_{K_{0}}\right\Vert=1$. Using $\text{tr}(AB)\leq\text{tr}(A^{\prime}A)^{1/2}\text{tr}(B^{\prime}B)^{1/2}$ for any two conformable matrices A and B, by the triangle inequality and Assumptions A.1(i), (iii) and (iv),

$$\begin{aligned}
\left|c_{K_0}' s_{2NT,4}\right| &= (NT)^{-3/2}\left|c_{K_0}'\sum_{i=1}^{N}X_{i(1)}'\sum_{k=1}^{K}\left(\beta_k^0-\tilde{\beta}_k^c\right)M_{F^0}\mathbf{X}_k'\sum_{l=1}^{K}\left(\beta_l^0-\tilde{\beta}_l^c\right)\mathbf{X}_l\tilde{F}_{(1)}\lambda_{i(1)}^*\right| \\
&= (NT)^{-3/2}\left|\mathrm{tr}\left(\tilde{F}_{(1)}\sum_{i=1}^{N}\lambda_{i(1)}^* c_{K_0}'X_{i(1)}'M_{F^0}\sum_{j=1}^{N}X_j\left(\beta^0-\tilde{\beta}^c\right)\left(\beta^0-\tilde{\beta}^c\right)'X_j'\right)\right| \\
&\leq (NT)^{-3/2}\varsigma_{1NT}^{1/2}\varsigma_{2NT}^{1/2},
\end{aligned}$$

where $\varsigma_{1NT}=\mathrm{tr}(\sum_{i=1}^{N}X_i(\beta^0-\tilde{\beta}^c)(\beta^0-\tilde{\beta}^c)'X_i'\;M_{F^0}\sum_{j=1}^{N}X_j(\beta^0-\tilde{\beta}^c)(\beta^0-\tilde{\beta}^c)'X_j')$ and $\varsigma_{2NT}=\mathrm{tr}(\tilde{F}_{(1)}\sum_{i=1}^{N}\lambda_{i(1)}^*c_{K_0}'X_{i(1)}'\sum_{j=1}^{N}X_{j(1)}c_{K_0}\lambda_{j(1)}^{*\prime}\tilde{F}_{(1)}')$. Using $\mu_{\max}(M_{F^0})=1$, the fact that $\mathrm{tr}(AB)\leq\mathrm{tr}(A)\mathrm{tr}(B)$ for any two conformable positive semidefinite matrix A and B, the rotational property of the trace operator, the fact that $A'BA\leq\mu_{\max}(B)A'A$ for any symmetric matrix B and conformable matrix A, and Assumptions A.1(i) and (viii), we have

$$\begin{aligned}
\varsigma_{1NT} &\leq \mathrm{tr}\left(\left(\beta^0-\tilde{\beta}^c\right)\left(\beta^0-\tilde{\beta}^c\right)'\sum_{i=1}^{N}\sum_{j=1}^{N}X_i'X_j\left(\beta^0-\tilde{\beta}^c\right)\left(\beta^0-\tilde{\beta}^c\right)'X_j'X_i\right) \\
&\leq \left\|\beta^0-\tilde{\beta}^c\right\|^2\mathrm{tr}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}X_i'X_j\left(\beta^0-\tilde{\beta}^c\right)\left(\beta^0-\tilde{\beta}^c\right)'X_j'X_i\right) \\
&= \left\|\beta^0-\tilde{\beta}^c\right\|^2\left(\beta^0-\tilde{\beta}^c\right)'\sum_{i=1}^{N}X_j'\sum_{j=1}^{N}X_iX_i'X_j\left(\beta^0-\tilde{\beta}^c\right) \\
&\leq \left\|\beta^0-\tilde{\beta}^c\right\|^4\mu_{\max}\left(\sum_{j=1}^{N}X_j'\sum_{i=1}^{N}X_iX_i'\;X_j\right)\leq\left\|\beta^0-\tilde{\beta}^c\right\|^4\mu_{\max}\left(\sum_{i=1}^{N}X_iX_i'\right)\mu_{\max}\left(\sum_{j=1}^{N}X_j'X_j\right) \\
&= \left\|\beta^0-\tilde{\beta}^c\right\|^4\mu_{\max}\left(\sum_{j=1}^{N}X_j'X_j\right)^2=O_P\left((NT/K)^{-2}\right)O_P\left((NT)^2\right)=O_P\left(K^2\right).
\end{aligned}$$

Similarly, noting that $\mu_{\max}\left(c_{K_0}c_{K_0}'\right)\leq\left\|c_{K_0}c_{K_0}'\right\|=1$, we have

$$\begin{aligned}
\varsigma_{2NT} &= \mathrm{tr}\left(\sum_{i=1}^{N}\lambda_{i(1)}^*c_{K_0}'X_{i(1)}'\sum_{j=1}^{N}X_{j(1)}c_{K_0}\lambda_{j(1)}^{*\prime}\tilde{F}_{(1)}'\tilde{F}_{(1)}\right)\leq\left\|\tilde{F}_{(1)}\right\|^2\sum_{i=1}^{N}\sum_{j=1}^{N}c_{K_0}'X_{i(1)}'X_{j(1)}c_{K_0}\lambda_{j(1)}^{*\prime}\lambda_{i(1)}^* \\
&\leq \left\|\tilde{F}_{(1)}\right\|^2\left\{\sum_{i=1}^{N}\sum_{j=1}^{N}c_{K_0}'X_{i(1)}'X_{j(1)}c_{K_0}c_{K_0}'X_{j(1)}'X_{i(1)}c_{K_0}\right\}^{1/2}\left\{\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\lambda_{j(1)}^{*\prime}\lambda_{i(1)}^*\right)^2\right\}^{1/2} \\
&\leq \left\|\tilde{F}_{(1)}\right\|^2\left\{c_{K_0}'\sum_{i=1}^{N}X_{i(1)}'\sum_{j=1}^{N}X_{j(1)}X_{j(1)}'X_{i(1)}c_{K_0}\right\}^{1/2}\sum_{i=1}^{N}\left\|\lambda_{i(1)}^*\right\|^2 \\
&\leq \left\|\tilde{F}_{(1)}\right\|^2\mu_{\max}\left(\sum_{i=1}^{N}X_{i(1)}'X_{i(1)}\right)\sum_{i=1}^{N}\left\|\lambda_{i(1)}^*\right\|^2=O_P(T)\,O_P(NT)\,O_P(N)=O_P\left(N^2T^2\right),
\end{aligned}$$

where the second inequality follows from the fact that $\mu_{\max}\left(\sum_{i=1}^{N}X_{i(1)}'X_{i(1)}\right)\leq\mu_{\max}\left(\sum_{i=1}^{N}X_i'X_i\right)=O_P(NT)$ under Assumption A.1(viii). It follows that $\left|c_{K_0}'s_{2NT,4}\right|=(NT)^{-3/2}O_P(K)\,O_P(NT)=o_P(1)$

under Assumption A.3(i), implying that $\|s_{2NT,4}\| = o_P(1)$. For $s_{2NT,7}$, we have

$$\begin{aligned}
\left|c_{K_0}' s_{2NT,7}\right| &= (NT)^{-3/2}\left|\sum_{i=1}^{N} c_{K_0}' X_{i(1)}' M_{F^0}\boldsymbol{\varepsilon}' \sum_{k=1}^{K}\left(\beta_k^0 - \tilde{\beta}_k^c\right)\mathbf{X}_k \tilde{F}_{(1)}\lambda_{i(1)}^*\right| \\
&= (NT)^{-3/2}\left|\operatorname{tr}\left(\tilde{F}_{(1)}\sum_{i=1}^{N}\lambda_{i(1)}^* c_{K_0}' X_{i(1)}' M_{F^0}\sum_{j=1}^{N}\varepsilon_j\left(\beta^0-\tilde{\beta}^c\right)' X_j'\right)\right| \le (NT)^{-3/2}\varsigma_{3NT}^{1/2}\varsigma_{2NT}^{1/2},
\end{aligned}$$

where $\varsigma_{3NT} = \operatorname{tr}(\sum_{i=1}^{N} X_i(\beta^0-\tilde{\beta}^c)\varepsilon_i' M_{F^0}\sum_{j=1}^{N}\varepsilon_j(\beta^0-\tilde{\beta}^c)' X_j')$. Note that

$$\begin{aligned}
\varsigma_{3NT} &\le \operatorname{tr}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} X_i\left(\beta^0-\tilde{\beta}^c\right)\varepsilon_i'\varepsilon_j\left(\beta^0-\tilde{\beta}^c\right)' X_j'\right) \\
&= \left(\beta^0-\tilde{\beta}^c\right)'\sum_{i=1}^{N}\sum_{j=1}^{N} X_j' X_i \varepsilon_i'\varepsilon_j\left(\beta^0-\tilde{\beta}^c\right) \le \left\|\beta^0-\tilde{\beta}^c\right\|^2 \mu_{\max}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} X_j' X_i\varepsilon_i'\varepsilon_j\right) \\
&= O_P\left((NT/K)^{-1}\right) O_P\left((N+T)NT\right) = O_P\left(K(N+T)\right),
\end{aligned}$$

where we use the fact that under Assumptions A.1(v) and (viii)

$$\begin{aligned}
\mu_{\max}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} X_j' X_i\varepsilon_i'\varepsilon_j\right) &= \max_{\varkappa=(\varkappa_1,\ldots,\varkappa_K)':\|\varkappa\|=1}\varkappa'\sum_{i=1}^{N}\sum_{j=1}^{N} X_j' X_i\varepsilon_i'\varepsilon_j\varkappa \\
&= \max_{\varkappa=(\varkappa_1,\ldots,\varkappa_K)':\|\varkappa\|=1}\operatorname{tr}\left(\sum_{m=1}^{K}\sum_{n=1}^{K}\varkappa_m\varkappa_n\mathbf{X}_m\mathbf{X}_n'\varepsilon\varepsilon'\right) \\
&\le \|\varepsilon\|_{\text{sp}}^2 \max_{\varkappa=(\varkappa_1,\ldots,\varkappa_K)':\|\varkappa\|=1}\sum_{m=1}^{K}\sum_{n=1}^{K}\varkappa_m\varkappa_n\operatorname{tr}\left(\mathbf{X}_m\mathbf{X}_n'\right) \\
&= \|\varepsilon\|_{\text{sp}}^2 \max_{\varkappa=(\varkappa_1,\ldots,\varkappa_K)':\|\varkappa\|=1}\sum_{m=1}^{K}\sum_{n=1}^{K}\sum_{i=1}^{N}\sum_{t=1}^{T}\varkappa_m\varkappa_n X_{it,m}X_{it,n} \\
&= \|\varepsilon\|_{\text{sp}}^2\mu_{\max}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}X_{it}'\right) = O_P(N+T)\,O_P(NT).
\end{aligned}$$

It follows that $\left|c_{K_0}' s_{2NT,7}\right| = (NT)^{-3/2} O_P(NT)\, O_P\left(K^{1/2}\left(N^{1/2}+T^{1/2}\right)\right) = O_P\left(K^{1/2}\left(N^{-1/2}+T^{-1/2}\right)\right)$ $= o_P(1)$, implying that $\|s_{2NT,7}\| = o_P(1)$. Analogously, we can show that $\|s_{2NT,8}\| = O_P(K^{1/2}(N^{-1/2}$ $+T^{-1/2})) = o_P(1)$.

For $s_{2NT,1}$, $s_{2NT,3}$, and $s_{2NT,6}$, we have $s_{2NT,1} = (NT)^{-3/2}\sum_{i=1}^{N} X_{i(1)}' M_{F^0}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\tilde{F}_{(1)}\lambda_{i(1)}^* = \mathbb{B}_{2nT}$, $s_{2NT,3} = (NT)^{-1/2}\sum_{i=1}^{N}\left[\frac{1}{NT}\sum_{j=1}^{N}\lambda_i^{0\prime}F^{0\prime}\tilde{F}_{(1)}\lambda_{j(1)}^* X_{j(1)}'\right] M_{F^0}\varepsilon_i \equiv \mathbb{V}_{1nT}$, and

$$\begin{aligned}
-s_{2NT,6} &= (NT)^{-3/2}\sum_{i=1}^{N} X_{i(1)}'\sum_{k=1}^{K}\left(\tilde{\beta}_k^c-\beta_k^0\right) M_{F^0}\mathbf{X}_k'\lambda^0 F^{0\prime}\tilde{F}_{(1)}\lambda_{i(1)}^* \\
&= \left[(NT)^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_j^{0\prime}F^{0\prime}\tilde{F}_{(1)}\lambda_{i(1)}^* X_{i(1)}' M_{F^0}X_j\right]\sqrt{NT}(\tilde{\beta}^c-\beta^0) \\
&= C_{NT}'\sqrt{NT}(\tilde{\beta}^c-\beta^0) \equiv \mathbb{V}_{2NT}.
\end{aligned}$$

Combining the above results yields the conclusion. ■

Proposition B.2 *Suppose that the conditions of Theorem 3.3 hold. Then* $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}X_{i(1)}'(M_{\hat{F}_{(1)}}-M_{F_{(1)}^{*}})\varepsilon_i=-\mathbb{B}_{3NT}+\mathbf{o}_P(1)$, *where* $\mathbb{B}_{3NT}$ *is defined in Section 3.3.*

Proof. Using (C.3) in the supplementary appendix, we have $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}X_{i(1)}'(M_{\hat{F}_{(1)}}-M_{F_{(1)}^{*}})\varepsilon_i=-\sum_{l=1}^{4}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}X_{i(1)}'b_l\varepsilon_i\equiv-\sum_{l=1}^{4}B_{lNT}$, say. Let $c_{K_0}=(c_{1K_0},...,c_{K_0K_0})'$ be an arbitrary $K_0\times 1$ nonrandom vector with $\|c_{K_0}\|=1$. Then by the facts that $|\mathrm{tr}(A)|\le\mathrm{rank}(A)\|A\|_{\mathrm{sp}}$ and $\|A\|_{\mathrm{sp}}\le\|A\|$, the submultiplicative property of the spectral norm, Lemmas A.2(i) and (v), and Assumptions A.1(iv)-(v) and A.3(i)

$$
\begin{aligned}
\left|c_{K_0}'B_{1NT}\right| &= (NT)^{-1/2}\left|\sum_{i=1}^{N}c_{K_0}'X_{i(1)}'\left(\hat{F}_{(1)}-F_{(1)}^{*}\right)\left(\hat{F}_{(1)}'\hat{F}_{(1)}\right)^{-1}\left(\hat{F}_{(1)}-F_{(1)}^{*}\right)'\varepsilon_i\right| \\
&= (NT)^{-1/2}\left|\mathrm{tr}\left[\left(\hat{F}_{(1)}-F_{(1)}^{*}\right)\left(\hat{F}_{(1)}'\hat{F}_{(1)}\right)^{-1}\left(\hat{F}_{(1)}-F_{(1)}^{*}\right)'\sum_{i=1}^{N}\varepsilon_i c_{K_0}'X_{i(1)}'\right]\right| \\
&\le R_0\,(NT)^{-1/2}\left\|\left(\frac{1}{T}\hat{F}_{(1)}'\hat{F}_{(1)}\right)^{-1}\right\|\frac{1}{T}\left\|\hat{F}_{(1)}-F_{(1)}^{*}\right\|^{2}\left\|\sum_{i=1}^{N}\varepsilon_i c_{K_0}'X_{i(1)}'\right\|_{\mathrm{sp}} \\
&= (NT)^{-1/2}O_P(1)\,O_P\left(\delta_{NT}^{-2}\right)O_P((N^{1/2}+T^{1/2})N^{1/2}T^{1/2})=o_P(1),
\end{aligned}
$$

where we use the fact that

$$
\begin{aligned}
\left\|\sum_{i=1}^{N}\varepsilon_i c_{K_0}'X_{i(1)}'\right\|_{\mathrm{sp}}^{2} &= \mu_{\max}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\varepsilon_i c_{K_0}'X_{i(1)}'X_{j(1)}c_{K_0}\varepsilon_j'\right)\le\|\boldsymbol{\varepsilon}\|_{\mathrm{sp}}^{2}\,\mu_{\max}\left(\sum_{i=1}^{N}X_{i(1)}'X_{i(1)}\right) \\
&= O_P(N+T)\,O_P(NT)\ \text{ by Assumptions A.1(v) and (viii).}
\end{aligned}
$$

By Lemma A.3(ii), $B_{2NT}=\mathbf{o}_P(1)$. By Lemma A.3(iii), $B_{3NT}=\mathbb{B}_{3NT}+\mathbf{o}_P(1)$. Lastly, by Lemma A.2(v) and Assumptions A.1(ii), (iv) and (vii)

$$
\begin{aligned}
\|B_{4NT}\| &= N^{-1/2}T^{-3/2}\left\|\sum_{i=1}^{N}X_{i(1)}'F_{(1)}^{*}\left[\left(T^{-1}\hat{F}_{(1)}'\hat{F}_{(1)}\right)^{-1}-\left(T^{-1}F_{(1)}^{*\prime}F_{(1)}^{*}\right)^{-1}\right]F_{(1)}^{*\prime}\varepsilon_i\right\| \\
&\le N^{-1/2}T^{-3/2}\left\|\left(T^{-1}\hat{F}_{(1)}'\hat{F}_{(1)}\right)^{-1}-\left(T^{-1}F_{(1)}^{*\prime}F_{(1)}^{*}\right)^{-1}\right\|\sum_{i=1}^{N}\left\|X_{i(1)}'F_{(1)}^{*}\right\|\left\|F_{(1)}^{*\prime}\varepsilon_i\right\| \\
&\le N^{-1/2}T^{-3/2}\left\|\left(T^{-1}\hat{F}_{(1)}'\hat{F}_{(1)}\right)^{-1}-\left(T^{-1}F_{(1)}^{*\prime}F_{(1)}^{*}\right)^{-1}\right\|\left\{\sum_{i=1}^{N}\left\|X_{i(1)}'F_{(1)}^{*}\right\|^{2}\right\}^{1/2}\left\{\sum_{i=1}^{N}\left\|F_{(1)}^{*\prime}\varepsilon_i\right\|^{2}\right\}^{1/2} \\
&= N^{-1/2}T^{-3/2}O_P\left(\delta_{NT}^{-2}+K(NT)^{-1/2}\right)O_P\left(N^{1/2}T\right)O_P\left((NT)^{1/2}\right) \\
&= O_P\left(N^{1/2}\delta_{NT}^{-2}+KT^{-1/2}\right)=o_P(1).
\end{aligned}
$$

Consequently, we have $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}X_{i(1)}'(M_{\hat{F}_{(1)}}-M_{F_{(1)}^{*}})\varepsilon_i=-\mathbb{B}_{3NT}+\mathbf{o}_P(1)$. ■

B.4 Proof of Theorem 3.5

By Theorem 3.2, we know that the shrinkage estimation based on γ_{NT}^{0} can correctly select all relevant covariates and factors and shrink the coefficients of irrelevant covariates and factors to 0 w.p.a.1. This implies that $\gamma_{NT}^0 \in \Omega_0$ and w.p.a.1

$$IC\left(\gamma_{NT}^0\right) = \hat{\sigma}^2\left(\gamma_{NT}^0\right) + \rho_{1NT}\left|\mathcal{S}_\beta\left(\gamma_{NT}^0\right)\right| + \rho_{2NT}N\left|\mathcal{S}_\lambda\left(\gamma_{NT}^0\right)\right| = \hat{\sigma}_{\mathcal{S}_T}^2 + o_P(1) \xrightarrow{P} \sigma_{\mathcal{S}_T}^2,$$

where the second equality holds by Theorem 3.2 and Assumption A.8, and the last convergence holds by Assumption A.7. We consider the cases of under- and over-fitted models separately.

Case 1: Under-fitted model. In this case, we have either $\mathcal{S}_\beta(\gamma) \not\supseteq \mathcal{S}_{T,\beta}$ or $\mathcal{S}_\lambda(\gamma) \not\supseteq \mathcal{S}_{T,\lambda}$. Noting that $\hat{\sigma}^2(\gamma) \geq \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(Y_{it} - \hat{\beta}_{\mathcal{S}_\beta(\gamma)}' X_{it,\mathcal{S}_\beta(\gamma)} - \hat{\lambda}_{i,\mathcal{S}_\lambda(\gamma)}'\hat{F}_{t,\mathcal{S}_\lambda(\gamma)})^2 = \hat{\sigma}_{\mathcal{S}(\gamma)}^2$, we have by Assumption A.7

$$\begin{aligned} IC(\gamma) &= \hat{\sigma}^2(\gamma) + \rho_{1NT}\left|\mathcal{S}_\beta(\gamma)\right| + \rho_{2NT}N\left|\mathcal{S}_\lambda(\gamma)\right| > \hat{\sigma}_{\mathcal{S}(\gamma)}^2 \\ &\geq \min_{\left\{\mathcal{S}:\ \mathcal{S}_\beta \not\supseteq \mathcal{S}_{T,\beta} \text{ or } \mathcal{S}_\lambda \not\supseteq \mathcal{S}_{T,\lambda}\right\}} \hat{\sigma}_{\mathcal{S}}^2 \xrightarrow{P} \min_{\left\{\mathcal{S}:\ \mathcal{S}_\beta \not\supseteq \mathcal{S}_{T,\beta} \text{ or } \mathcal{S}_\lambda \not\supseteq \mathcal{S}_{T,\lambda}\right\}} \sigma_{\mathcal{S}}^2 > \sigma_{\mathcal{S}_T}^2. \end{aligned}$$

It follows that $P\left(\inf_{\gamma\in\Omega_-} IC(\gamma) > IC\left(\gamma_{NT}^0\right)\right) \to 1$.

Case 2: Over-fitted model. Let $\mathbb{S} = \{\mathcal{S}_\beta \times \mathcal{S}_\lambda : \mathcal{S}_\beta \supset \mathcal{S}_{T,\beta},\ \mathcal{S}_\lambda \supset \mathcal{S}_{T,\lambda},\ \ |\mathcal{S}_\beta| + |\mathcal{S}_\lambda| > |\mathcal{S}_{T,\beta}| + |\mathcal{S}_{T,\lambda}|\}$. Let $\gamma \in \Omega_+$ such that $\mathcal{S}(\gamma) = \mathcal{S}_\beta(\gamma) \times \mathcal{S}_\lambda(\gamma) \in \mathbb{S}$. Let

$$L_{NT,\mathcal{S}}\left(\beta_{\mathcal{S}_\beta}, \lambda_{\mathcal{S}_\lambda}\right) = \frac{1}{NT}\sum_{i=1}^N\left\|Y_i - X_{i,\mathcal{S}_\beta}\beta_{\mathcal{S}_\beta} - \hat{F}_{\mathcal{S}_\lambda}\lambda_{i,\mathcal{S}_\lambda}\right\|^2, \tag{B.5}$$

where $\mathcal{S} = \mathcal{S}_\beta \times \mathcal{S}_\lambda$, $\lambda_{\mathcal{S}_\lambda} = (\lambda_{1,\mathcal{S}_\lambda}, ..., \lambda_{N,\mathcal{S}_\lambda})'$, $X_{i,\mathcal{S}_\beta}$ denotes the $T \times |\mathcal{S}_\beta|$ submatrix of X_i with column indices given by $\mathcal{S}_\beta$, and $\hat{F}_{\mathcal{S}_\lambda}$ is analogously defined. By the definition of $\hat{\sigma}_{\mathcal{S}}^2$ in (3.6), we have $\hat{\sigma}_{\mathcal{S}}^2 = L_{NT,\mathcal{S}}(\hat{\beta}_{\mathcal{S}_\beta}, \hat{\lambda}_{\mathcal{S}_\lambda})$. In view of the facts that $P\left(\hat{\sigma}^2\left(\gamma_{NT}^0\right) = \hat{\sigma}_{\mathcal{S}_T}^2\right) \to 1$ as $(N,T) \to \infty$ by Theorem 3.2 and that $\hat{\sigma}^2(\gamma) \geq \hat{\sigma}_{\mathcal{S}(\gamma)}^2$, we have that w.p.a.1

$$\begin{aligned} &\delta_{NT}^2[IC(\gamma) - IC\left(\gamma_{NT}^0\right)] \\ =\ &\delta_{NT}^2[\hat{\sigma}^2(\gamma) - \hat{\sigma}^2\left(\gamma_{NT}^0\right)] + \delta_{NT}^2\left[\rho_{1NT}\left(|\mathcal{S}_\beta(\gamma)| - K_0\right) + \rho_{2NT}N\left(|\mathcal{S}_\lambda(\gamma)| - R_0\right)\right] \\ \geq\ &\delta_{NT}^2[\hat{\sigma}_{\mathcal{S}(\gamma)}^2 - \hat{\sigma}_{\mathcal{S}_T}^2] + \delta_{NT}^2\left[\rho_{1NT}\left(|\mathcal{S}_\beta(\gamma)| - K_0\right) + \rho_{2NT}N\left(|\mathcal{S}_\lambda(\gamma)| - R_0\right)\right]. \end{aligned}$$

By Assumption A.8, $\delta_{NT}^2\rho_{1NT} \to \infty$ and $\delta_{NT}^2\rho_{2NT}N \to \infty$. In addition, for any $\mathcal{S} = \mathcal{S}_\beta \times \mathcal{S}_\lambda \in \mathbb{S}$ we have $\max(|\mathcal{S}_\beta| - K_0,\ |\mathcal{S}_\lambda| - R_0) \geq 1$ and $\min(|\mathcal{S}_\beta| - K_0,\ |\mathcal{S}_\lambda| - R_0) \geq 0$. Then by Proposition B.3 below

$$\begin{aligned} &P\left(\inf_{\gamma\in\Omega_+} IC(\gamma) > IC\left(\gamma_{NT}^0\right)\right) \\ =\ &P\left(\inf_{\gamma\in\Omega_+}\delta_{NT}^2[IC(\gamma) - IC\left(\gamma_{NT}^0\right)] > 0\right) \\ \geq\ &P\left(\min_{\mathcal{S}=\mathcal{S}_\beta\times\mathcal{S}_\lambda\in\mathbb{S}}\delta_{NT}^2[\hat{\sigma}_{\mathcal{S}}^2 - \hat{\sigma}_{\mathcal{S}_T}^2] + \delta_{NT}^2\left[\rho_{1NT}\left(|\mathcal{S}_\beta| - K_0\right) + \rho_{2NT}N\left(|\mathcal{S}_\lambda| - R_0\right)\right] > 0\right) \\ \to\ &1 \text{ as } (N,T) \to \infty.\ \blacksquare \end{aligned}$$

Proposition B.3 *Suppose that the conditions in Theorem 3.5 hold. Suppose that* $\mathcal{S} \in \mathbb{S} = \{\mathcal{S}_\beta \times \mathcal{S}_\lambda : \mathcal{S}_\beta \supset \mathcal{S}_{T,\beta},\ \mathcal{S}_\lambda \supset \mathcal{S}_{T,\lambda},\ |\mathcal{S}_\beta| + |\mathcal{S}_\lambda| > |\mathcal{S}_{T,\beta}| + |\mathcal{S}_{T,\lambda}|\}$. *Then* $\hat{\sigma}_{\mathcal{S}}^2 - \hat{\sigma}_{\mathcal{S}_T}^2 = O_P(\delta_{NT}^{-2})$.

Proof. Consider the minimization of the following objective function

$$\bar{L}_{NT,\mathcal{S}_T}\left(\beta_{\mathcal{S}_{T,\beta}}, \lambda\right) = \frac{1}{NT}\sum_{i=1}^N\left\|Y_i - X_{i,\mathcal{S}_{T,\beta}}\beta_{\mathcal{S}_{T,\beta}} - F^0\lambda_i\right\|^2, \tag{B.6}$$

where we pretend that the factors are observed. Let $\bar{\beta}_{\mathcal{S}_{T,\beta}}$ and $\bar{\lambda}$ denote the OLS solution to the above problem. Let $\bar{\sigma}^2_{\mathcal{S}_T} \equiv \bar{L}_{NT,\mathcal{S}_T}\left(\bar{\beta}_{\mathcal{S}_{T,\beta}}, \bar{\lambda}\right)$. Straightforward algebra shows that

$$\bar{\sigma}^2_{\mathcal{S}_T} = \frac{1}{NT}\sum_{i=1}^{N}\left(Y_i - X_{i,\mathcal{S}_{T,\beta}}\bar{\beta}_{\mathcal{S}_{T,\beta}}\right)' M_{F^0}\left(Y_i - X_{i,\mathcal{S}_{T,\beta}}\bar{\beta}_{\mathcal{S}_{T,\beta}}\right) \tag{B.7}$$

and

$$\left\|\bar{\beta}_{\mathcal{S}_{T,\beta}} - \beta^0_{\mathcal{S}_{T,\beta}}\right\| = O_P\left((NT/K_0)^{-1/2}\right). \tag{B.8}$$

Noting that $|\hat{\sigma}^2_{\mathcal{S}} - \hat{\sigma}^2_{\mathcal{S}_T}| \leq |\hat{\sigma}^2_{\mathcal{S}} - \bar{\sigma}^2_{\mathcal{S}_T}| + |\hat{\sigma}^2_{\mathcal{S}_T} - \bar{\sigma}^2_{\mathcal{S}_T}| \leq 2\max_{\mathcal{S}_\beta \supset \mathcal{S}_{T,\beta},\, \mathcal{S}_\lambda \supset \mathcal{S}_{T,\lambda}} |\hat{\sigma}^2_{\mathcal{S}} - \bar{\sigma}^2_{\mathcal{S}_T}|$, it suffices to prove that for each $\mathcal{S} \in \bar{\mathbb{S}} \equiv \mathbb{S} \cup \mathcal{S}_T$ with $\mathcal{S}_T = \mathcal{S}_{T,\beta} \times \mathcal{S}_{T,\lambda}$, we have $\hat{\sigma}^2_{\mathcal{S}} - \bar{\sigma}^2_{\mathcal{S}_T} = O_P(\delta^{-2}_{NT})$.

Now, fix $\mathcal{S} \in \bar{\mathbb{S}}$. We consider the minimization of the least square objective function in (B.5). Noting that $T^{-1}\hat{F}'_{\mathcal{S}_\lambda}\hat{F}_{\mathcal{S}_\lambda}$ is asymptotically singular when $|\mathcal{S}_\lambda| > R_0$, the OLS estimate $\hat{\lambda}_{\mathcal{S}_\lambda(\gamma)}$ is not necessarily unique whereas $\hat{\beta}_{\mathcal{S}_\beta(\gamma)}$ is. Despite this, the minimum of $L_{NT,\mathcal{S}}(\beta_{\mathcal{S}_\beta}, \lambda_{\mathcal{S}_\lambda})$ is uniquely determined. Standard algebra shows that $\hat{\beta}_{\mathcal{S}_\beta} = \left(\frac{1}{NT}\sum_{i=1}^N X'_{i,\mathcal{S}_\beta} M_{\hat{F}_{\mathcal{S}_\lambda}} X_{i,\mathcal{S}_\beta}\right)^{-1} \frac{1}{NT}\sum_{i=1}^N X'_{i,\mathcal{S}_\beta} M_{\hat{F}_{\mathcal{S}_\lambda}} Y_i$, and $\hat{\lambda}_{i,\mathcal{S}_\lambda} = (\hat{F}'_{\mathcal{S}_\lambda}\hat{F}_{\mathcal{S}_\lambda})^{-}\hat{F}'_{\mathcal{S}_\lambda}(Y_i - X_{i,\mathcal{S}_\beta}\hat{\beta}_{\mathcal{S}_\beta})$, where A^- is any generalized inverse of A and $M_{\hat{F}_{\mathcal{S}_\lambda}} = I_T - \hat{F}_{\mathcal{S}_\lambda}(\hat{F}'_{\mathcal{S}_\lambda}\hat{F}_{\mathcal{S}_\lambda})^{+}\hat{F}'_{\mathcal{S}_\lambda}$. It follows that

$$\hat{\sigma}^2_{\mathcal{S}} = \frac{1}{NT}L_{NT,\mathcal{S}}\left(\hat{\beta}_{\mathcal{S}_\beta}, \hat{\lambda}_{\mathcal{S}_\lambda}\right) = \frac{1}{NT}\sum_{i=1}^{N}\left(Y_i - X_{i,\mathcal{S}_\beta}\hat{\beta}_{\mathcal{S}_\beta}\right)' M_{\hat{F}_{\mathcal{S}_\lambda}}\left(Y_i - X_{i,\mathcal{S}_\beta}\hat{\beta}_{\mathcal{S}_\beta}\right). \tag{B.9}$$

In addition, using Lemma A.3 we can readily show that

$$\left\|\hat{\beta}_{\mathcal{S}_\beta} - \beta^0_{\mathcal{S}_\beta}\right\| = O_P\left(K^{1/2}\delta^{-2}_{NT}\right). \tag{B.10}$$

Recall $F^* = F^0 H$ and $\lambda^*_i = H^{+}\lambda^0_i$. Define $F^*_{\mathcal{S}_\lambda}$ as an $N \times |\mathcal{S}_\lambda|$ submatrix of F^* whose column indices are given in $\mathcal{S}_\lambda$. Similarly define $\lambda^*_{i,\mathcal{S}_\lambda}$. Noting that $Y_i - X_{i,\mathcal{S}_\beta}\hat{\beta}_{\mathcal{S}_\beta} = (X_{i,\mathcal{S}_\beta}\beta^0_{\mathcal{S}_\beta} + F^*_{\mathcal{S}_\lambda}\lambda^*_{i,\mathcal{S}_\lambda} + \varepsilon_i) - X_{i,\mathcal{S}_\beta}\hat{\beta}_{\mathcal{S}_\beta} = \hat{F}_{\mathcal{S}_\lambda}\lambda^*_{i,\mathcal{S}_\lambda} + e_i$ where $e_i = \varepsilon_i + X_{i,\mathcal{S}_\beta}(\beta^0_{\mathcal{S}_\beta} - \hat{\beta}_{\mathcal{S}_\beta}) + (F^*_{\mathcal{S}_\lambda} - \hat{F}_{\mathcal{S}_\lambda})\lambda^*_{i,\mathcal{S}_\lambda}$, we can decompose $\hat{\sigma}^2_{\mathcal{S}} = \frac{1}{NT}\sum_{i=1}^N e'_i M_{\hat{F}_{\mathcal{S}_\lambda}} e_i$ as follows

$$\begin{aligned}
\hat{\sigma}^2_{\mathcal{S}} &= \frac{1}{NT}\sum_{i=1}^{N}\varepsilon'_i M_{\hat{F}_{\mathcal{S}_\lambda}}\varepsilon_i + \frac{1}{NT}\sum_{i=1}^{N}(\beta^0_{\mathcal{S}_\beta} - \hat{\beta}_{\mathcal{S}_\beta})' X'_{i,\mathcal{S}_\beta} M_{\hat{F}_{\mathcal{S}_\lambda}} X_{i,\mathcal{S}_\beta}(\beta^0_{\mathcal{S}_\beta} - \hat{\beta}_{\mathcal{S}_\beta}) \\
&\quad + \frac{1}{NT}\sum_{i=1}^{N}\lambda^{*\prime}_{i,\mathcal{S}_\lambda}(F^*_{\mathcal{S}_\lambda} - \hat{F}_{\mathcal{S}_\lambda})' M_{\hat{F}_{\mathcal{S}_\lambda}}(F^*_{\mathcal{S}_\lambda} - \hat{F}_{\mathcal{S}_\lambda})\lambda^*_{i,\mathcal{S}_\lambda} + \frac{2}{NT}\sum_{i=1}^{N}\varepsilon'_i M_{\hat{F}_{\mathcal{S}_\lambda}} X_{i,\mathcal{S}_\beta}(\beta^0_{\mathcal{S}_\beta} - \hat{\beta}_{\mathcal{S}_\beta}) \\
&\quad + \frac{2}{NT}\sum_{i=1}^{N}\varepsilon'_i M_{\hat{F}_{\mathcal{S}_\lambda}}(F^*_{\mathcal{S}_\lambda} - \hat{F}_{\mathcal{S}_\lambda})\lambda^*_{i,\mathcal{S}_\lambda} + \frac{2}{NT}\sum_{i=1}^{N}(\beta^0_{\mathcal{S}_\beta} - \hat{\beta}_{\mathcal{S}_\beta}) X'_{i,\mathcal{S}_\beta} M_{\hat{F}_{\mathcal{S}_\lambda}}(F^*_{\mathcal{S}_\lambda} - \hat{F}_{\mathcal{S}_\lambda})\lambda^*_{i,\mathcal{S}_\lambda} \\
&\equiv \hat{\Delta}_{1NT} + \hat{\Delta}_{2NT} + \hat{\Delta}_{3NT} + 2\hat{\Delta}_{4NT} + 2\hat{\Delta}_{5NT} + 2\hat{\Delta}_{6NT}, \text{ say.}
\end{aligned}$$

By (B.10), Lemma A.2(i), Assumption A.1(viii), and the fact that $\mu_{\max}(M_{\hat{F}_{\mathcal{S}_\lambda}}) = 1$, $\hat{\Delta}_{2NT} \leq ||\beta^0_{\mathcal{S}_\beta} - \hat{\beta}_{\mathcal{S}_\beta}||^2 \mu_{\max}(\frac{1}{NT}\sum_{i=1}^N X'_{i,\mathcal{S}_\beta} X_{i,\mathcal{S}_\beta}) = O_P(K\delta^{-4}_{NT})$, and $\hat{\Delta}_{3NT} \leq \frac{1}{T}||F^*_{\mathcal{S}_\lambda} - \hat{F}_{\mathcal{S}_\lambda}||^2 \frac{1}{N}\sum_{i=1}^N \left\|\lambda^*_{i,\mathcal{S}_\lambda}\right\|^2 = O_P(\delta^{-2}_{NT})$.

By CS inequality, $|\hat{\Delta}_{6NT}| \leq \{\hat{\Delta}_{2NT}\hat{\Delta}_{3NT}\}^{1/2} = O_P(K^{1/2}\delta_{NT}^{-3})$. For $\hat{\Delta}_{5NT}$, we have

$$\begin{aligned}
\left|\hat{\Delta}_{5NT}\right| &= \frac{1}{NT}\left|\operatorname{tr}\left(M_{\hat{F}_{\mathcal{S}_\lambda}}(F^*_{\mathcal{S}_\lambda}-\hat{F}_{\mathcal{S}_\lambda})\sum_{i=1}^N \lambda^*_{i,\mathcal{S}_\lambda}\varepsilon_i'\right)\right| \\
&\leq \frac{|\mathcal{S}_\lambda|}{\sqrt{N}}\frac{1}{\sqrt{T}}\left\|F^*_{\mathcal{S}_\lambda}-\hat{F}_{\mathcal{S}_\lambda}\right\|\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^N \lambda^*_{i,\mathcal{S}_\lambda}\varepsilon_i'\right\| = O_P\left(N^{-1/2}\delta_{NT}^{-1}\right)
\end{aligned}$$

by Lemma A.2(i), the fact that $|\operatorname{tr}(A)| \leq r\|A\|$ for any $r\times r$ matrix A, the submultiplicative property of the Frobenius norm, and the facts that $\mu_{\max}(M_{\hat{F}_{\mathcal{S}_\lambda}}) = 1$ and that $\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^N \lambda^*_{i,\mathcal{S}_\lambda}\varepsilon_i'\right\| = O_P(1)$. Next, by the triangle inequality

$$\begin{aligned}
\left|\hat{\Delta}_{4NT}\right| &\leq \frac{1}{NT}\left|\sum_{i=1}^N \varepsilon_i' X_{i,\mathcal{S}_\beta}(\beta^0_{\mathcal{S}_\beta}-\hat{\beta}_{\mathcal{S}_\beta})\right| + \frac{1}{NT}\left|\sum_{i=1}^N \varepsilon_i' P_{\hat{F}_{\mathcal{S}_\lambda}} X_{i,\mathcal{S}_\beta}(\beta^0_{\mathcal{S}_\beta}-\hat{\beta}_{\mathcal{S}_\beta})\right| \\
&\equiv \hat{\Delta}_{4NT,1}+\hat{\Delta}_{4NT,2}, \text{ say.}
\end{aligned}$$

By (B.10) and the fact that $\left\|\frac{1}{NT}\sum_{i=1}^N \varepsilon_i' X_{i,\mathcal{S}_\beta}\right\| = O_P\left((NT/K)^{-1/2}\right)$ under Assumption A.1(vi), we have

$$\hat{\Delta}_{4NT,1} \leq \left\|\frac{1}{NT}\sum_{i=1}^N \varepsilon_i' X_{i,\mathcal{S}_\beta}\right\|\left\|\beta^0_{\mathcal{S}_\beta}-\hat{\beta}_{\mathcal{S}_\beta}\right\| = O_P\left(K(NT)^{-1/2}\delta_{NT}^{-2}\right).$$

By CS inequality, Lemma A.2(i) and Assumption A.1(v)

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^N\left\|\hat{F}'_{\mathcal{S}_\lambda}\varepsilon_i\right\|^2 &\leq \frac{2}{NT}\sum_{i=1}^N\left\|H^{0\prime}_{\mathcal{S}_\lambda}F^{0\prime}\varepsilon_i\right\|^2 + \frac{2}{NT}\sum_{i=1}^N\left\|\left(\hat{F}_{\mathcal{S}_\lambda}-F^0H_{\mathcal{S}_\lambda}\right)'\varepsilon_i\right\|^2 \\
&\leq \left\|H^0_{\mathcal{S}_\lambda}\right\|^2\frac{2}{NT}\sum_{i=1}^N\left\|F^{0\prime}\varepsilon_i\right\|^2 + \frac{2}{NT}\operatorname{tr}\left[\left(\hat{F}_{\mathcal{S}_\lambda}-F^0H_{\mathcal{S}_\lambda}\right)'\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\left(\hat{F}_{\mathcal{S}_\lambda}-F^0H_{\mathcal{S}_\lambda}\right)\right] \\
&= O_P(1) + N^{-1}T^{-1}\left\|\hat{F}_{\mathcal{S}_\lambda}-F^0H_{\mathcal{S}_\lambda}\right\|^2\|\boldsymbol{\varepsilon}\|^2_{\text{sp}} = O_P(1).
\end{aligned}$$

This result, in conjunction with CS inequality and (B.10), yields

$$\begin{aligned}
\hat{\Delta}_{4NT,2} &= \left\{\frac{1}{NT}\sum_{i=1}^N \varepsilon_i' P_{\hat{F}_{\mathcal{S}_\lambda}}\varepsilon_i\right\}^{1/2}\left\{(\beta^0_{\mathcal{S}_\beta}-\hat{\beta}_{\mathcal{S}_\beta})\frac{1}{NT}\sum_{i=1}^N X'_{i,\mathcal{S}_\beta}X_{i,\mathcal{S}_\beta}(\beta^0_{\mathcal{S}_\beta}-\hat{\beta}_{\mathcal{S}_\beta})\right\}^{1/2} \\
&\leq \left\|\left(T^{-1}\hat{F}'_{\mathcal{S}_\lambda}\hat{F}_{\mathcal{S}_\lambda}\right)^+\right\|^{1/2}\left\{\frac{1}{NT^2}\sum_{i=1}^N\left\|\hat{F}'_{\mathcal{S}_\lambda}\varepsilon_i\right\|^2\right\}^{1/2}\left\|\beta^0_{\mathcal{S}_\beta}-\hat{\beta}_{\mathcal{S}_\beta}\right\|\upsilon_{2NT}^{1/2} \\
&= O_P\left(K^{1/2}T^{-1/2}\delta_{NT}^{-2}\right),
\end{aligned}$$

where $\upsilon_{2NT} = \mu_{\max}(\frac{1}{NT}\sum_{i=1}^N X'_{i,\mathcal{S}_\beta}X_{i,\mathcal{S}_\beta}) = O_P(1)$ under Assumption A.1(viii). Thus $\hat{\Delta}_{4NT} = o_P\left(\delta_{NT}^{-2}\right)$. Consequently, we have shown that $\hat{\sigma}^2_{\mathcal{S}} = \frac{1}{NT}\sum_{i=1}^N \varepsilon_i' M_{\hat{F}_{\mathcal{S}_\lambda}}\varepsilon_i + O_P(\delta_{NT}^{-2})$.

Analogously, using $Y_i - X_{i,\mathcal{S}_{T,\beta}}\bar{\beta}_{\mathcal{S}_{T,\beta}} = F^0\lambda_i^0 + \varepsilon_i + X_{i,\mathcal{S}_{T,\beta}}(\beta^0_{\mathcal{S}_{T,\beta}} - \bar{\beta}_{\mathcal{S}_{T,\beta}})$, (B.7), and (B.8) we can readily show that $\bar{\sigma}^2_{\mathcal{S}_T} = \frac{1}{NT}\sum_{i=1}^N \varepsilon_i' M_{F^0}\varepsilon_i + O_P((NT)^{-1})$. It follows that $\hat{\sigma}^2_{\mathcal{S}} - \bar{\sigma}^2_{\mathcal{S}_T} = \frac{1}{NT}\sum_{i=1}^N(\varepsilon_i' P_{F^0}\varepsilon_i - \varepsilon_i' P_{\hat{F}_{\mathcal{S}_\lambda}}\varepsilon_i) + O_P(\delta_{NT}^{-2})$. Noting that $\frac{1}{NT}\sum_{i=1}^N \varepsilon_i' P_{F^0}\varepsilon_i = O_P\left(T^{-1}\right)$, we are left to show that $R_{NT} \equiv \frac{1}{NT}\sum_{i=1}^N \varepsilon_i' P_{\hat{F}_{\mathcal{S}_\lambda}}\varepsilon_i = O_P(\delta_{NT}^{-2})$. Note that $R_{NT} = \frac{1}{NT}\operatorname{tr}(P_{\hat{F}_{\mathcal{S}_\lambda}}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}) \leq \frac{1}{NT}\|\boldsymbol{\varepsilon}\|^2_{sp}\operatorname{tr}(P_{\hat{F}_{\mathcal{S}_\lambda}}) = \frac{|\mathcal{S}_\lambda|}{NT}\|\boldsymbol{\varepsilon}\|^2_{\text{sp}} = O_P(\delta_{NT}^{-2})$. Consequently, we have shown that $\hat{\sigma}^2_{\mathcal{S}} - \bar{\sigma}^2_{\mathcal{S}_T} = O_P(\delta_{NT}^{-2})$ for any $\mathcal{S} \in \bar{\mathbb{S}}$. ■

REFERENCES

Ahn, S., Horenstein, A., 2013. Eigenvalue ratio test for the number of factors. Econometrica 81, 1203-1227.

Bai, J., 2003. Inferential theory for factors models of large dimensions. Econometrica 71, 135-173.

Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica 77, 1229-1279.

Bai, J., Liao, Y., 2013. Efficient estimation of approximate factor models via regularized maximum likelihood. Working paper, University of Maryland.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191-221.

Bai, J., Ng, S., 2006. Evaluating latent and observed factors in macroeconomics and finance. Journal of Econometrics 131, 507-537.

Bai, J., Ng, S., 2007. Determining the number of primitive shocks in factor models. Journal of Business & Economic Statistics 25, 52-60.

Barro, R., 1991. Economic growth in a cross section of countries. Quarterly Journal of Economics 106, 407-43.

Belloni, A., Chernozhukov, V., 2013. Least squares after model selection in high-dimensional sparse models. Bernoulli 19, 521-547.

Belloni, A., Chernozhukov, V., Hansen, C., 2012. Sparse models and methods for optimal instruments with an application to eminent domain. Econometrica 80, 2369-2429.

Bernstein, D. S., 2005. Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory. Princeton University Press, Princeton.

Caner, M., 2009. Lasso-type GMM Estimator. Econometric Theory 25, 270-290.

Caner, M., Han, X., 2014. Selecting the correct number of factors in approximate factor models: the large panel case with Bridge estimators. Journal of Business & Economic Statistics 32, 359-374.

Caner, M., Knight, K., 2013. An alternative to unit root tests: Bridge estimators differentiate between nonstationary versus stationary models and select optimal lag. Journal of Statistical Planning and Inference 143, 691-715.

Cheng, X., Liao, Z., 2015. Select the valid and relevant moments: an information-based LASSO for GMM with many moments. Journal of Econometrics, forthcoming.

Cheng, X., Liao, Z., Schorfheide F., 2014. Shrinkage estimation of high-dimensional factor models with structural instabilities. Working paper, University of Pennsylvania.

Cochrane, D., Orcutt, G. H., 1949. Application of least squares regression to relationships containing auto-correlated error terms. Journal of the American Statistical Association 44, 32-61.

Durlauf, S., Johnson, P., Temple, J., 2005. Growth econometrics. In P. Aghion and S. N. Durlauf (eds.), Handbook of Economic Growth, Volume 1A, North-Holland, Amsterdam.

Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 18, 61-90.

Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association 96, 1348-1360.

Fan, J., Liao, Y., 2014. Endogeneity in high dimensions. Annals of Statistics 42, 872-917.

Fan, J., Peng, H., 2004. Nonconcave penalized likelihood with a diverging number of parameters. Annals of Statistics 32, 928-961.

García, P. E. 2011. Instrumental variable estimation and selection with many weak and irrelevant instruments. Working paper, University of Wisconsin, Madison.

Greenaway-McGrevy, R., C. Han, D. Sul, 2012. Asymptotic distribution of factor augmented estimators for panel regression. Journal of Econometrics 169, 48-53.

Hallin, M., Liška, R., 2007. Determining the number of factors in the general dynamic factor model. Journal of the American Statistical Association 102, 603-617.

Hirose, K., Konishi, S., 2012. Variable selection via the weighted group lasso for factor analysis models. Canadian Journal of Statistics 40, 345-361.

Huang, J., Horowitz, J. L., Ma, S., 2008. Asymptotic properties of bridge estimators in sparse high-dimensional regression models. Annals of Statistics 36, 587-613.

Islam, N., 1995. Growth empirics: a panel data approach. Quarterly Journal of Economics 110, 1127-1170.

Kapetanios, G., 2010. A testing procedure for determining the number of factors in approximate factor models with large datasets. Journal of Business & Economic Statistics 28, 397-409.

Kapetanios, G., Pesaran, M. H., 2007. Alternative approaches to estimation and inference in large multifactor panels: small sample results with an application to modelling of asset return. In G. Phillips and E. Tzavalis (eds.), The Refinement of Econometric Estimation and Testing Procedures: Finite Sample and Asymptotic Analysis, Cambridge University Press, Cambridge.

Knight, K., Fu, W., 2000. Asymptotics for Lasso-type estimators. Annals of Statistics 28, 1356-1378.

Kock, A. B., 2013. Oracle efficient variable selection in random and fixed effects panel data models. Econometric Theory 29, 115-152.

Lam, C., Fan, J., 2008. Profile-kernel likelihood inference with diverging number of parameters. Annals of Statistics 36, 2232-2260.

Lee, N., Moon, H. R., Weidner, M., 2012. Analysis of interactive fixed effects dynamic linear panel regression with measurement error. Economics Letters 117, 239-242.

Leeb, H., Pötscher, P. M., 2005. Model selection and inference: facts and fiction. Econometric Theory 21, 21-59.

Leeb, H., Pötscher, P. M., 2008. Sparse estimators and the oracle property, or the return of Hodges' estimator. Journal of Econometrics 142, 201-211.

Leeb, H., Pötscher, P. M., 2009. On the distribution of penalized maximum likelihood estimators: the LASSO, SCAD, and thresholding. Journal of Multivariate Analysis 100, 2065-2082.

Liao, Z., 2013. Adaptive GMM shrinkage estimation with consistent moment selection. Econometric Theory 29, 857-904.

Liao, Z., Phillips, P. C. B., 2015. Automated estimation of vector error correction models. Econometric Theory, forthcoming.

Lu, X., Su, L., 2015. Jackknife model averaging for quantile regressions. Journal of Econometrics, forthcoming.

Moon, H., Weidner, M., 2014a. Linear regression for panel with unknown number of factors as interactive fixed effects. Working paper, University of Southern California.

Moon, H., Weidner, M., 2014b. Dynamic linear panel regression models with interactive fixed effects. Working paper, University of Southern California.

Moon, H., Shum, M., Weidner, M., 2014. Estimation of random coefficients logit demand models with interactive fixed effects. Working paper, University of Southern California.

Moral-Benito, E., 2012. Determinants of economic growth: a Bayesian panel data approach. Review of Economics and Statistics, 94, 566–579.

Onatski, A., 2009. Testing hypotheses about the number of factors in large factor models. Econometrica 77, 1447-1479.

Onatski, A., 2010. Determining the number of factors from empirical distribution of eigenvalues. Review of Economics and Statistics 92, 1004-1016.

Onatski, A., 2012. Asymptotics of the principal components estimator of large factor models with weakly influential factors. Journal of Econometrics 168, 244-258.

Pesaran, M. H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74, 967-1012.

Pesaran, M. H., Tosetti, E., 2011. Large panels with common factors and spatial correlation. Journal of Econometrics 161, 182-202.

Prakasa Rao, B. L. S., 2009. Conditional independence, conditional mixing and conditional association. Annals of the Institute of Statistical Mathematics 61, 441-460.

Roussas, G. G., 2008. On conditional independence, mixing and association. Stochastic Analysis and Applications 26, 1274-1309.

Sala-i-Martin, X., Doppelhofer, G., Miller, R., 2004. Determinants of long-term growth: a Bayesian averaging of classical estimates (BACE) approach. American Economic Review 94, 813-835.

Schneider, U., Pötscher, P. M., 2009. On the distribution of the adaptive LASSO estimator. Journal of Statistical Planning and Inference 139, 2775-2790.

Su, L., Chen, Q., 2013. Testing homogeneity in panel data models with interactive fixed effects. Econometric Theory 29, 1079-1135.

Su, L., Jin, S., 2012. Sieve estimation of panel data models with cross section dependence. Journal of Econometrics 169, 34-47.

Su, L., Jin, S., Zhang, Y., 2015. Specification test for panel data models with interactive fixed effects. Journal of Econometrics, forthcoming.

Su, L., Zhang, Y., 2014. Nonparametric dynamic panel data models with interactive fixed effects: sieve estimation and specification testing. Working paper, Singapore Management University.

Tibshirani, R. J., 1996. Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society, Series B. 58, 267-288.

Wang, H., Leng, C., 2008. A note of adaptive group Lasso. Computational Statistics and Data Analysis 52, 5277-5286.

Wang, H., Li, R., Tsai, C-L., 2007. Tuning parameter selectors for the smoothly clipped absolute deviation method. Biometrika 94, 553-568.

Wei, F., Huang, J., 2010. Consistent group selection in high-dimensional linear regression. Bernoulli 16, 1369-1384.

Yuan, M., Lin, Y., 2006. Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society, Series B. 68, 49-67.

Zou, H., 2006. The adaptive Lasso and its oracle properties. Journal of the American Statistical Association 101, 1418-1429.

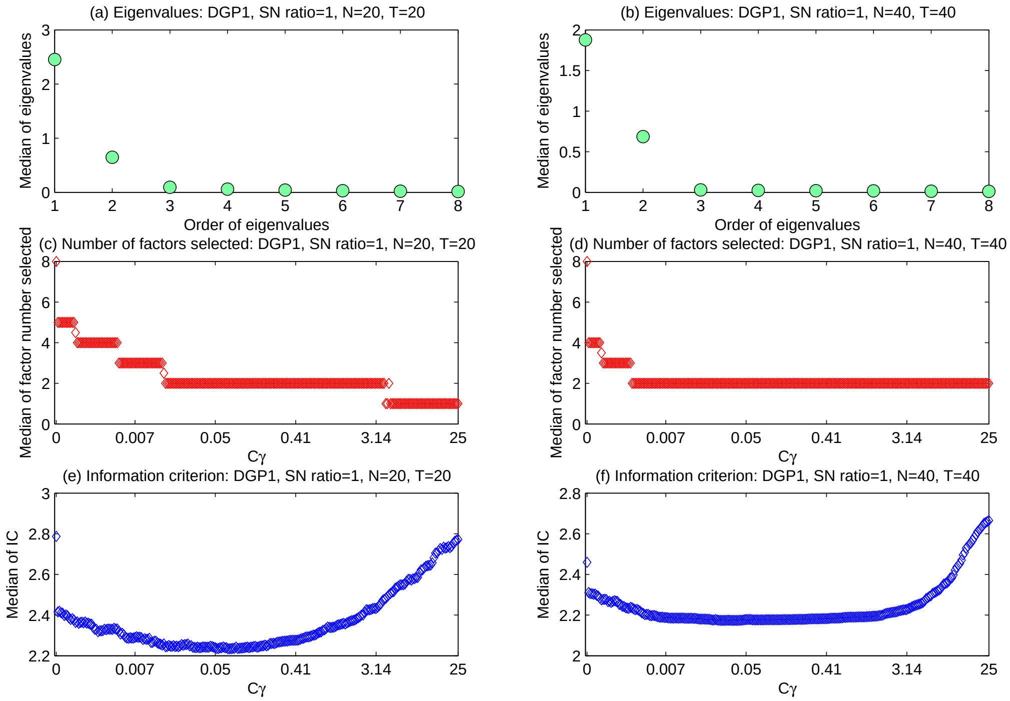

Figure 1: Illustration of the main ideas

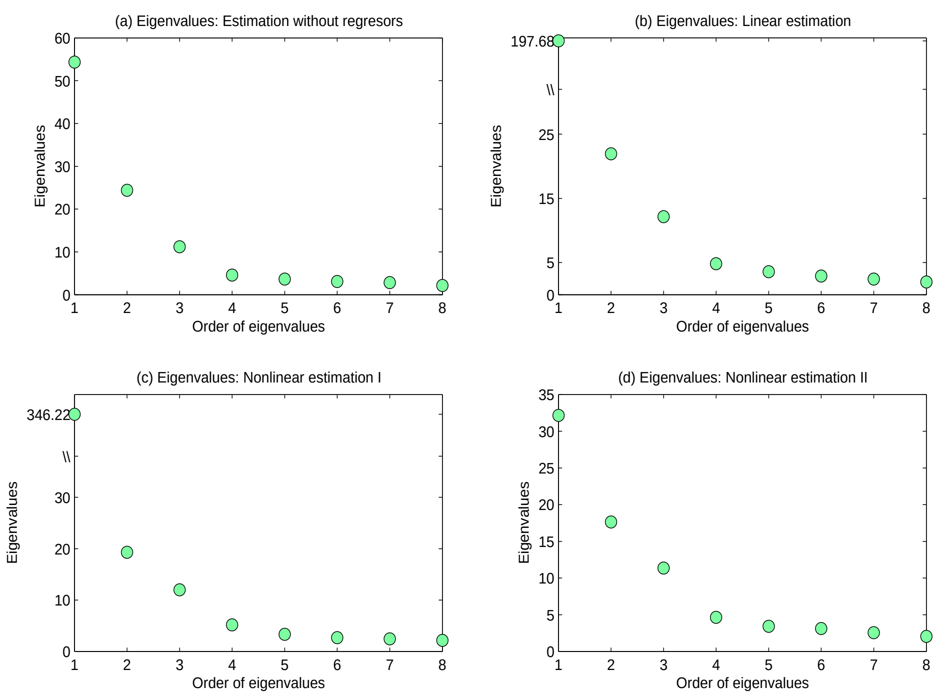

Figure 2: Eigenvalues used in estimation

Table 1: MSEs of estimates of (β_1, β_2) for different numbers of factors included in the model

				Number of factors included					
DGP	N	T		0	1	**2**	4	6	8
2	20	20	β_1	19.37	6.47	**3.77**	4.43	6.49	9.96
			β_2	20.56	6.61	**3.84**	3.71	5.73	9.67
	40	40	β_1	18.25	4.68	**0.57**	0.62	0.75	0.92
			β_2	17.80	4.21	**0.42**	0.48	0.66	0.85
	20	60	β_1	17.13	3.99	**0.73**	0.80	1.08	1.57
			β_2	18.67	4.69	**0.80**	0.91	1.23	1.71
	60	20	β_1	18.62	4.25	**0.72**	0.82	1.11	1.52
			β_2	19.46	4.86	**0.88**	0.92	1.19	1.67
	60	60	β_1	16.75	3.81	**0.23**	0.25	0.29	0.31
			β_2	16.71	3.89	**0.19**	0.21	0.24	0.26
3	20	20	β_1	30.40	8.74	**4.24**	4.87	6.52	11.32
			β_2	1.45	0.63	**0.44**	1.46	5.47	14.70
	40	40	β_1	26.59	5.82	**0.59**	0.74	0.86	0.97
			β_2	1.35	0.45	**0.06**	0.11	0.23	0.67
	20	60	β_1	28.09	6.30	**0.87**	1.11	1.25	1.75
			β_2	1.40	0.48	**0.09**	0.17	0.26	0.44
	60	20	β_1	27.10	6.06	**0.86**	1.07	1.80	6.18
			β_2	1.31	0.36	**0.12**	0.79	4.88	15.48
	60	60	β_1	27.06	6.13	**0.23**	0.28	0.33	0.39
			β_2	1.34	0.40	**0.03**	0.04	0.06	0.10
4	20	20	β_1	11.71	6.17	**4.15**	5.19	7.61	10.57
			β_2	11.83	6.24	**4.02**	3.98	6.32	11.05
	40	40	β_1	10.13	3.58	**0.58**	0.73	0.85	0.96
			β_2	10.26	3.67	**0.54**	0.65	0.73	0.89
	20	60	β_1	10.05	3.74	**0.98**	1.02	1.32	1.78
			β_2	10.11	3.70	**0.88**	0.97	1.23	1.68
	60	20	β_1	11.00	4.00	**0.95**	1.08	1.35	1.77
			β_2	10.63	3.81	**0.70**	0.90	1.04	1.39
	60	60	β_1	9.71	2.97	**0.19**	0.21	0.25	0.27
			β_2	9.80	2.99	**0.20**	0.22	0.25	0.29
5	20	20	β_1	19.24	8.22	**4.67**	4.94	7.68	18.05
			β_2	0.97	0.60	**0.53**	2.20	9.58	33.48
	40	40	β_1	16.51	4.81	**0.60**	0.73	0.86	1.10
			β_2	0.80	0.34	**0.06**	0.11	0.27	1.23
	20	60	β_1	17.57	5.32	**0.91**	1.12	1.26	1.75
			β_2	0.85	0.39	**0.11**	0.19	0.28	0.46
	60	20	β_1	16.97	5.08	**0.93**	1.11	4.74	17.77
			β_2	0.80	0.28	**0.13**	1.56	12.15	37.73
	60	60	β_1	16.90	4.91	**0.24**	0.28	0.33	0.39
			β_2	0.78	0.29	**0.03**	0.04	0.06	0.12
6	20	20	β_1	6.77	5.88	**4.62**	4.80	6.98	9.72
			β_2	6.97	5.53	**4.33**	4.68	7.05	11.18
	40	40	β_1	5.29	2.86	**0.46**	0.56	0.64	0.89
			β_2	5.64	3.05	**0.57**	0.65	0.75	0.89
	20	60	β_1	5.54	3.26	**0.87**	0.90	1.22	1.55
			β_2	5.42	3.12	**0.68**	0.82	1.11	1.38
	60	20	β_1	5.95	3.40	**0.82**	1.15	1.44	1.84
			β_2	5.61	3.03	**0.80**	0.94	1.21	1.62
	60	60	β_1	5.00	2.21	**0.18**	0.21	0.23	0.26
			β_2	5.01	2.20	**0.21**	0.24	0.29	0.32

Note: Numbers in the main entries are 100×MSEs of the estimates of β_1 or β_2.

Table 2: Selection of the number of factors (SN ratio=1)

				Comparison methods								AgLasso			
				Bai and Ng-				Ona-	Ona-	AH-		IC	Rule of thumb		
DGP	N	T		PC_1	PC_2	IC_1	IC_2	ReSt	Eca	ER	GR		c=0.5	c=1	c=2
1	20	20	$r<2$	0.00	0.00	0.02	0.14	0.08	0.71	0.33	0.23	0.00	0.17	0.28	0.42
			$\mathbf{r=2}$	**0.00**	**0.00**	**0.46**	**0.86**	**0.87**	**0.22**	**0.67**	**0.77**	**0.76**	**0.82**	**0.72**	**0.57**
			$r>2$	1.00	1.00	0.52	0.00	0.05	0.07	0.00	0.00	0.24	0.01	0.00	0.00
	40	40	$r<2$	0.00	0.00	0.00	0.00	0.00	0.07	0.02	0.01	0.00	0.02	0.03	0.07
			$\mathbf{r=2}$	**0.00**	**0.72**	**1.00**	**1.00**	**1.00**	**0.88**	**0.98**	**0.99**	**0.98**	**0.98**	**0.97**	**0.93**
			$r>2$	1.00	0.28	0.00	0.00	0.00	0.06	0.00	0.00	0.02	0.00	0.00	0.00
	20	60	$r<2$	0.00	0.00	0.01	0.01	0.00	0.20	0.10	0.06	0.00	0.04	0.09	0.20
			$\mathbf{r=2}$	**0.00**	**0.00**	**0.99**	**0.99**	**1.00**	**0.73**	**0.90**	**0.94**	**1.00**	**0.96**	**0.91**	**0.80**
			$r>2$	1.00	1.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00
	60	20	$r<2$	0.00	0.00	0.00	0.01	0.00	0.26	0.09	0.07	0.00	0.08	0.10	0.17
			$\mathbf{r=2}$	**0.00**	**0.00**	**1.00**	**0.99**	**0.98**	**0.69**	**0.91**	**0.93**	**1.00**	**0.92**	**0.90**	**0.83**
			$r>2$	1.00	1.00	0.00	0.00	0.02	0.05	0.00	0.00	0.00	0.00	0.00	0.00
	60	60	$r<2$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
			$\mathbf{r=2}$	**1.00**	**1.00**	**1.00**	**1.00**	**0.99**	**0.95**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**0.99**
			$r>2$	0.00	0.00	0.00	0.00	0.01	0.05	0.00	0.00	0.00	0.00	0.00	0.00
2	20	20	$r<2$	0.00	0.00	0.29	0.92	0.74	0.92	0.56	0.55	0.16	0.04	0.13	0.30
			$\mathbf{r=2}$	**0.00**	**0.00**	**0.14**	**0.08**	**0.22**	**0.01**	**0.23**	**0.25**	**0.54**	**0.37**	**0.49**	**0.47**
			$r>2$	1.00	1.00	0.57	0.00	0.04	0.07	0.21	0.20	0.31	0.59	0.38	0.23
	40	40	$r<2$	0.00	0.00	0.31	0.70	0.10	0.80	0.30	0.24	0.00	0.01	0.05	0.11
			$\mathbf{r=2}$	**0.02**	**0.69**	**0.69**	**0.30**	**0.90**	**0.16**	**0.70**	**0.76**	**0.94**	**0.72**	**0.85**	**0.86**
			$r>2$	0.98	0.31	0.00	0.00	0.01	0.04	0.00	0.00	0.06	0.27	0.10	0.03
	20	60	$r<2$	0.00	0.00	0.52	0.70	0.32	0.85	0.43	0.39	0.34	0.03	0.06	0.14
			$\mathbf{r=2}$	**0.00**	**0.00**	**0.48**	**0.30**	**0.67**	**0.08**	**0.56**	**0.59**	**0.65**	**0.47**	**0.67**	**0.76**
			$r>2$	1.00	1.00	0.00	0.00	0.01	0.07	0.01	0.02	0.00	0.50	0.27	0.10
	60	20	$r<2$	0.00	0.00	0.51	0.69	0.30	0.90	0.51	0.46	0.33	0.03	0.09	0.18
			$\mathbf{r=2}$	**0.00**	**0.00**	**0.49**	**0.31**	**0.69**	**0.08**	**0.49**	**0.54**	**0.67**	**0.55**	**0.70**	**0.76**
			$r>2$	1.00	1.00	0.00	0.00	0.01	0.03	0.00	0.00	0.00	0.42	0.20	0.06
	60	60	$r<2$	0.00	0.01	0.06	0.28	0.00	0.44	0.08	0.04	0.00	0.00	0.00	0.04
			$\mathbf{r=2}$	**0.99**	**0.99**	**0.94**	**0.72**	**0.98**	**0.52**	**0.92**	**0.96**	**1.00**	**0.95**	**0.99**	**0.96**
			$r>2$	0.01	0.00	0.00	0.00	0.02	0.04	0.00	0.00	0.00	0.05	0.00	0.00
3	20	20	$r<2$	0.00	0.00	0.06	0.50	0.52	0.88	0.56	0.53	0.08	0.17	0.24	0.36
			$\mathbf{r=2}$	**0.00**	**0.00**	**0.03**	**0.37**	**0.36**	**0.04**	**0.32**	**0.33**	**0.42**	**0.47**	**0.56**	**0.56**
			$r>2$	1.00	1.00	0.91	0.14	0.11	0.08	0.12	0.14	0.49	0.36	0.20	0.08
	40	40	$r<2$	0.00	0.00	0.16	0.45	0.07	0.69	0.30	0.24	0.00	0.01	0.06	0.13
			$\mathbf{r=2}$	**0.00**	**0.62**	**0.84**	**0.55**	**0.92**	**0.23**	**0.70**	**0.76**	**0.90**	**0.82**	**0.88**	**0.87**
			$r>2$	1.00	0.38	0.00	0.00	0.01	0.08	0.00	0.00	0.10	0.17	0.06	0.00
	20	60	$r<2$	0.00	0.00	0.34	0.52	0.15	0.78	0.38	0.34	0.16	0.03	0.08	0.18
			$\mathbf{r=2}$	**0.00**	**0.00**	**0.66**	**0.48**	**0.83**	**0.16**	**0.62**	**0.66**	**0.84**	**0.67**	**0.77**	**0.78**
			$r>2$	1.00	1.00	0.00	0.00	0.02	0.07	0.00	0.00	0.00	0.30	0.15	0.04
	60	20	$r<2$	0.00	0.00	0.04	0.12	0.19	0.86	0.50	0.41	0.19	0.08	0.16	0.28
			$\mathbf{r=2}$	**0.00**	**0.00**	**0.54**	**0.81**	**0.80**	**0.07**	**0.50**	**0.59**	**0.80**	**0.77**	**0.77**	**0.70**
			$r>2$	1.00	1.00	0.42	0.08	0.02	0.06	0.00	0.00	0.01	0.15	0.06	0.02
	60	60	$r<2$	0.00	0.00	0.03	0.15	0.00	0.29	0.09	0.06	0.00	0.01	0.03	0.05
			$\mathbf{r=2}$	**1.00**	**1.00**	**0.97**	**0.85**	**0.99**	**0.64**	**0.91**	**0.94**	**1.00**	**0.99**	**0.97**	**0.95**
			$r>2$	0.00	0.00	0.00	0.00	0.01	0.07	0.00	0.00	0.00	0.00	0.00	0.00

Notes: Numbers in the main entries are the proportions of the replications in which the selected number of factors is less than, equal to, or greater than the true number of factors (i.e., 2) out of total 250 replications. Bai and Ng refers to Bai and Ng (2002), Ona-ReSt refers to Onatski (2010), Ona-Eca refers to Onatski (2009) and AH refers to Ahn and Horenstein (2013).

Table 2: Selection of the number of factors (SN ratio=1) (cont'd)

				Comparison methods								AgLasso			
				Bai and Ng-				Ona-	Ona-	AH-		IC	Rule of thumb		
DGP	N	T		PC_1	PC_2	IC_1	IC_2	ReSt	Eca	ER	GR		c=0.5	c=1	c=2
4	20	20	$r < 2$	0.00	0.00	0.19	0.93	0.75	0.94	0.48	0.50	0.24	0.06	0.12	0.25
			$\mathbf{r = 2}$	**0.00**	**0.00**	**0.04**	**0.07**	**0.20**	**0.02**	**0.21**	**0.24**	**0.44**	**0.26**	**0.38**	**0.45**
			$r > 2$	1.00	1.00	0.77	0.00	0.05	0.04	0.30	0.26	0.32	0.68	0.50	0.30
	40	40	$r < 2$	0.00	0.00	0.33	0.72	0.10	0.82	0.27	0.23	0.00	0.01	0.04	0.12
			$\mathbf{r = 2}$	**0.00**	**0.66**	**0.67**	**0.28**	**0.90**	**0.12**	**0.73**	**0.77**	**0.97**	**0.67**	**0.82**	**0.84**
			$r > 2$	1.00	0.34	0.00	0.00	0.00	0.06	0.00	0.00	0.02	0.32	0.15	0.04
	20	60	$r < 2$	0.00	0.00	0.54	0.76	0.32	0.83	0.46	0.42	0.36	0.02	0.04	0.12
			$\mathbf{r = 2}$	**0.00**	**0.00**	**0.46**	**0.24**	**0.67**	**0.10**	**0.52**	**0.57**	**0.64**	**0.40**	**0.67**	**0.76**
			$r > 2$	1.00	1.00	0.00	0.00	0.01	0.07	0.02	0.02	0.00	0.58	0.29	0.12
	60	20	$r < 2$	0.00	0.00	0.49	0.68	0.32	0.88	0.44	0.39	0.34	0.02	0.06	0.14
			$\mathbf{r = 2}$	**0.00**	**0.00**	**0.51**	**0.32**	**0.67**	**0.09**	**0.56**	**0.60**	**0.65**	**0.54**	**0.69**	**0.73**
			$r > 2$	1.00	1.00	0.00	0.00	0.01	0.03	0.01	0.01	0.01	0.45	0.24	0.13
	60	60	$r < 2$	0.00	0.00	0.05	0.32	0.00	0.38	0.07	0.04	0.00	0.00	0.01	0.03
			$\mathbf{r = 2}$	**1.00**	**1.00**	**0.95**	**0.68**	**1.00**	**0.58**	**0.93**	**0.96**	**1.00**	**0.96**	**0.98**	**0.97**
			$r > 2$	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.04	0.01	0.00
5	20	20	$r < 2$	0.00	0.00	0.01	0.20	0.48	0.88	0.52	0.46	0.13	0.14	0.23	0.34
			$\mathbf{r = 2}$	**0.00**	**0.00**	**0.00**	**0.11**	**0.34**	**0.03**	**0.30**	**0.31**	**0.32**	**0.53**	**0.57**	**0.57**
			$r > 2$	1.00	1.00	0.99	0.70	0.18	0.09	0.18	0.23	0.55	0.34	0.20	0.09
	40	40	$r < 2$	0.00	0.00	0.16	0.42	0.07	0.73	0.31	0.24	0.00	0.02	0.05	0.13
			$\mathbf{r = 2}$	**0.00**	**0.44**	**0.84**	**0.58**	**0.92**	**0.21**	**0.69**	**0.76**	**0.90**	**0.83**	**0.88**	**0.86**
			$r > 2$	1.00	0.56	0.00	0.00	0.01	0.06	0.00	0.00	0.10	0.15	0.06	0.01
	20	60	$r < 2$	0.00	0.00	0.35	0.51	0.18	0.79	0.41	0.36	0.23	0.04	0.08	0.15
			$\mathbf{r = 2}$	**0.00**	**0.00**	**0.65**	**0.49**	**0.80**	**0.17**	**0.59**	**0.64**	**0.77**	**0.63**	**0.75**	**0.79**
			$r > 2$	1.00	1.00	0.00	0.00	0.02	0.04	0.00	0.00	0.00	0.34	0.17	0.06
	60	20	$r < 2$	0.00	0.00	0.00	0.00	0.30	0.89	0.52	0.42	0.34	0.08	0.18	0.29
			$\mathbf{r = 2}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.56**	**0.03**	**0.45**	**0.49**	**0.56**	**0.68**	**0.70**	**0.65**
			$r > 2$	1.00	1.00	1.00	1.00	0.13	0.08	0.04	0.08	0.10	0.24	0.12	0.06
	60	60	$r < 2$	0.00	0.00	0.04	0.15	0.00	0.30	0.08	0.06	0.00	0.01	0.03	0.06
			$\mathbf{r = 2}$	**1.00**	**1.00**	**0.96**	**0.85**	**0.99**	**0.63**	**0.92**	**0.94**	**1.00**	**0.99**	**0.97**	**0.94**
			$r > 2$	0.00	0.00	0.00	0.00	0.01	0.07	0.00	0.00	0.00	0.00	0.00	0.00
6	20	20	$r < 2$	0.00	0.00	0.05	0.88	0.75	0.95	0.46	0.47	0.22	0.04	0.10	0.20
			$\mathbf{r = 2}$	**0.00**	**0.00**	**0.01**	**0.11**	**0.17**	**0.02**	**0.17**	**0.20**	**0.51**	**0.28**	**0.41**	**0.50**
			$r > 2$	1.00	1.00	0.94	0.01	0.08	0.04	0.37	0.33	0.27	0.67	0.49	0.31
	40	40	$r < 2$	0.00	0.00	0.29	0.67	0.09	0.77	0.26	0.21	0.01	0.01	0.04	0.08
			$\mathbf{r = 2}$	**0.01**	**0.52**	**0.71**	**0.33**	**0.89**	**0.18**	**0.73**	**0.78**	**0.94**	**0.66**	**0.82**	**0.86**
			$r > 2$	0.99	0.48	0.00	0.00	0.02	0.06	0.00	0.00	0.05	0.33	0.14	0.06
	20	60	$r < 2$	0.00	0.00	0.48	0.66	0.27	0.86	0.39	0.36	0.32	0.01	0.04	0.11
			$\mathbf{r = 2}$	**0.00**	**0.00**	**0.52**	**0.34**	**0.71**	**0.11**	**0.60**	**0.62**	**0.68**	**0.46**	**0.64**	**0.76**
			$r > 2$	1.00	1.00	0.00	0.00	0.02	0.03	0.01	0.01	0.00	0.53	0.32	0.13
	60	20	$r < 2$	0.00	0.00	0.44	0.64	0.30	0.86	0.45	0.40	0.34	0.02	0.07	0.15
			$\mathbf{r = 2}$	**0.00**	**0.00**	**0.56**	**0.36**	**0.69**	**0.08**	**0.54**	**0.60**	**0.65**	**0.48**	**0.70**	**0.73**
			$r > 2$	1.00	1.00	0.00	0.00	0.01	0.06	0.01	0.01	0.01	0.49	0.24	0.12
	60	60	$r < 2$	0.00	0.00	0.05	0.30	0.00	0.39	0.05	0.04	0.00	0.00	0.00	0.02
			$\mathbf{r = 2}$	**1.00**	**1.00**	**0.95**	**0.70**	**0.99**	**0.54**	**0.95**	**0.96**	**1.00**	**0.95**	**0.99**	**0.97**
			$r > 2$	0.00	0.00	0.00	0.00	0.01	0.07	0.00	0.00	0.00	0.05	0.01	0.00

Table 3: Selection of regressors (SN ratio=1)

DGP	N	T	β_1	β_2	β_3	β_4	β_5
4	20	20	0.00	0.00	0.80	0.74	0.72
	40	40	0.00	0.00	0.98	0.98	0.98
	20	60	0.00	0.00	0.94	0.97	0.93
	60	20	0.00	0.00	0.94	0.95	0.92
	60	60	0.00	0.00	1.00	1.00	1.00
5	20	20	0.00	0.22	0.70	0.67	0.10
	40	40	0.00	0.00	0.94	0.92	0.70
	20	60	0.00	0.00	0.92	0.94	0.90
	60	20	0.00	0.11	0.76	0.78	0.04
	60	60	0.00	0.00	1.00	1.00	0.98
6	20	20	0.00	0.00	0.84	0.87	0.87
	40	40	0.00	0.00	0.99	1.00	0.98
	20	60	0.00	0.00	0.99	0.98	0.99
	60	20	0.00	0.00	0.98	0.98	0.98
	60	60	0.00	0.00	1.00	1.00	1.00

Note: Numbers in the main entries are the proportions of the replications in which the estimates of β's are shrunk to zeros out of total 250 replications.

Table 4: MSEs of the estimates of $\beta_1$and β_2 (SN ratio=1)

				Comparison methods								AgLasso		
				Bai and Ng-				Ona-	Ona-	AH-			BC-	Post
DGP	N	T		PC_1	PC_2	IC_1	IC_2	ReSt	Eca	ER	GR	AgLasso	AgLasso	AgLasso
2	20	20	β_1	9.98	9.40	7.94	11.18	7.94	17.53	5.07	4.72	4.08	4.11	3.83
			β_2	9.69	9.02	7.70	11.33	7.84	19.15	5.48	5.46	4.02	4.08	3.79
	40	40	β_1	0.63	0.58	1.43	4.33	0.90	13.96	1.21	1.04	0.63	0.65	0.57
			β_2	0.49	0.42	1.31	4.20	0.80	13.41	1.13	0.96	0.45	0.47	0.40
	20	60	β_1	1.24	1.01	2.30	3.56	1.70	13.68	1.47	1.33	1.35	1.45	1.31
			β_2	1.47	1.18	2.66	4.56	1.91	14.89	1.80	1.65	1.71	1.83	1.65
	60	20	β_1	1.21	1.03	1.95	3.53	1.45	15.77	1.81	1.65	1.39	1.42	1.33
			β_2	1.22	1.12	2.41	4.05	1.68	16.36	2.02	1.89	1.66	1.70	1.60
	60	60	β_1	0.23	0.23	0.34	1.14	0.23	6.76	0.40	0.31	0.23	0.23	0.23
			β_2	0.19	0.19	0.28	1.11	0.19	6.56	0.33	0.26	0.19	0.19	0.19
3	20	20	β_1	11.33	10.73	10.53	7.48	7.03	23.10	5.83	5.65	3.87	3.97	4.30
			β_2	14.67	13.41	13.96	1.73	0.78	1.42	0.92	0.99	5.00	3.11	0.95
	40	40	β_1	0.75	0.61	1.09	2.99	0.84	17.86	1.55	1.30	0.62	0.64	0.62
			β_2	0.13	0.07	0.07	0.18	0.07	0.89	0.11	0.08	0.25	0.16	0.07
	20	60	β_1	1.35	1.21	2.32	3.88	1.71	21.03	2.13	1.94	1.45	1.53	1.39
			β_2	0.28	0.24	0.17	0.26	0.13	1.04	0.16	0.15	0.17	0.16	0.12
	60	20	β_1	5.77	4.65	2.49	1.12	1.54	21.82	2.45	2.05	1.62	1.95	1.31
			β_2	14.65	12.05	4.02	0.17	0.14	1.28	0.13	0.15	2.18	1.09	0.12
	60	60	β_1	0.23	0.23	0.30	0.63	0.23	6.75	0.46	0.33	0.24	0.24	0.23
			β_2	0.03	0.03	0.03	0.05	0.03	0.36	0.04	0.03	0.06	0.04	0.03
4	20	20	β_1	10.57	10.61	10.07	7.73	6.28	10.53	6.67	6.08	5.06	5.07	4.37
			β_2	11.05	10.73	9.67	7.82	6.39	10.72	6.27	5.78	4.46	4.46	3.97
	40	40	β_1	0.72	0.63	1.31	3.51	0.84	8.05	1.06	0.96	0.62	0.64	0.58
			β_2	0.66	0.57	1.44	3.51	0.82	7.88	1.05	0.99	0.58	0.59	0.55
	20	60	β_1	1.39	1.26	2.41	3.65	2.03	8.43	1.91	1.77	1.82	1.92	1.82
			β_2	1.39	1.27	2.28	3.34	1.84	8.38	1.77	1.60	1.78	1.89	1.82
	60	20	β_1	1.46	1.28	1.96	3.13	1.70	9.24	1.71	1.57	1.60	1.64	1.61
			β_2	1.13	0.96	1.83	2.86	1.35	8.85	1.40	1.25	1.26	1.29	1.27
	60	60	β_1	0.19	0.19	0.26	0.83	0.19	3.54	0.29	0.24	0.19	0.19	0.18
			β_2	0.20	0.20	0.27	0.86	0.20	3.34	0.31	0.24	0.20	0.20	0.20
5	20	20	β_1	18.05	17.89	17.90	15.25	8.59	15.51	6.60	7.13	4.82	4.70	4.70
			β_2	33.48	33.35	33.44	26.14	4.63	1.12	3.21	4.79	8.65	5.68	3.00
	40	40	β_1	0.78	0.65	1.04	2.47	0.82	11.42	1.45	1.19	0.65	0.68	0.61
			β_2	0.17	0.09	0.07	0.15	0.07	0.58	0.09	0.08	0.49	0.33	0.09
	20	60	β_1	1.42	1.27	2.12	3.31	1.78	13.42	2.18	1.96	1.74	1.85	1.65
			β_2	0.30	0.25	0.17	0.23	0.15	0.66	0.17	0.16	0.21	0.18	0.14
	60	20	β_1	17.77	17.75	17.68	17.22	3.21	13.81	2.69	2.72	3.58	4.31	2.05
			β_2	37.73	37.69	37.54	36.58	3.78	1.12	0.90	1.58	2.92	1.70	1.05
	60	60	β_1	0.24	0.25	0.31	0.57	0.24	4.54	0.41	0.34	0.24	0.25	0.23
			β_2	0.03	0.03	0.03	0.04	0.03	0.22	0.03	0.03	0.07	0.05	0.03
6	20	20	β_1	9.72	9.84	9.55	5.73	5.36	6.67	6.42	6.14	4.38	4.44	4.52
			β_2	11.18	11.20	10.92	5.53	5.36	6.93	6.58	5.78	5.00	5.01	4.81
	40	40	β_1	0.55	0.49	1.06	2.24	0.61	4.11	0.87	0.73	0.49	0.49	0.45
			β_2	0.67	0.56	1.13	2.39	0.69	4.26	0.88	0.79	0.59	0.59	0.54
	20	60	β_1	1.33	1.19	1.73	2.27	1.19	4.75	1.41	1.34	1.36	1.42	1.39
			β_2	1.16	1.09	1.54	2.40	1.11	4.52	1.25	1.19	1.22	1.29	1.25
	60	20	β_1	1.64	1.47	1.62	2.17	1.26	4.98	1.36	1.30	1.44	1.47	1.42
			β_2	1.41	1.24	1.53	2.06	1.29	4.76	1.48	1.35	1.48	1.50	1.46
	60	60	β_1	0.18	0.18	0.23	0.68	0.18	1.96	0.24	0.22	0.19	0.19	0.19
			β_2	0.21	0.21	0.26	0.70	0.21	1.99	0.25	0.25	0.21	0.21	0.21

Notes: Numbers in the main entries are 100×MSEs of the estimates of β_1 or β_2. Bai and Ng refers to Bai and Ng (2002), Ona-ReSt refers to Onatski (2010), Ona-Eca refers to Onatski (2009) and AH refers to Ahn and Horenstein (2013).

Table 5: Summary statistics

Variables	Description	Mean	Median	SD	Min	Max	Data sources
			Dependent variable:				
Growth	Annual growth rate of real GDP per capita	1.57	1.84	6.12	-70.89	76.75	Penn Table
			Independent variables:				
Young	Age dependency ratio, young (% of working-age population)	66.02	72.92	23.15	19.34	106.43	WDI
Fert	Fertility rate (births per woman)	4.27	4.25	1.98	0.90	8.29	WDI
Life	Life expectancy at birth (years)	62.21	63.98	11.78	26.82	82.03	WDI
Popu	Population growth	1.92	2.05	1.38	-17.28	17.91	Penn Table
Invpri	Price level of investment	88.36	62.33	174.75	9.88	2612.60	Penn Table
Con	Consumption share	71.77	71.08	17.89	8.64	193.96	Penn Table
Gov	Government consumption share	10.54	8.40	7.63	0.73	58.64	Penn Table
Inv	Investment share	22.42	21.24	10.43	-11.50	80.12	Penn Table
Open	Openness	59.56	51.97	37.21	3.78	377.79	Penn Table

Table 6: The number of factors determined by various methods

	Bai and Ng			AH		
	IC_1	IC_2	Ona-ReSt	ER	GR	AgLasso
Estimation without regressors	3	2	3	3	3	3
Linear estimation	3	3	3	1	1	3
Nonlinear estimation I	3	3	3	1	1	3
Nonlinear estimation II	3	3	3	3	3	3

Note: Bai and Ng refers to Bai and Ng (2002), Ona-ReSt refers to Onatski (2010), and AH refers to Ahn and Horenstein (2013).

Table 7: Linear estimation

	Young	Fert	Life	Popu	Invpri	Con	Gov	Inv	Open	Lag1	Lag2	Lag3
						Number of factors=0						
estimate	0.018	-0.001	0.019	-0.446	0.001	-0.026	-0.053	0.085	0.002	0.143	0.041	0.031
t-stat	0.987	-0.003	1.754	-1.103	1.215	-2.907*	-2.604*	5.598*	0.732	2.904*	1.521	1.160
						Number of factors=3						
estimate	0.026	-0.680	-0.011	-0.161	0.002	-0.083	-0.159	0.146	-0.011	0.060	0.017	0.035
t-stat	1.241	-2.382*	-0.390	-1.081	1.717	-5.013*	-4.153*	6.195*	-1.842	2.094*	0.753	1.706
						Number of factors=5						
estimate	0.006	0.090	0.005	-0.441	0.000	-0.009	-0.157	0.264	-0.004	0.067	-0.064	-0.023
t-stat	0.127	0.159	0.148	-0.881	0.058	-0.319	-2.089*	5.002*	-0.360	1.277	-1.428	-0.501
						Number of factors=8						
estimate	-0.001	-0.040	-0.033	-0.234	0.003	0.003	-0.048	0.323	-0.006	0.092	-0.061	-0.046
t-stat	-0.032	-0.094	-1.379	-1.209	1.461	0.153	-0.875	9.215*	-0.894	2.652*	-2.052*	-1.507
						AgLasso: Number of factors=3						
estimate	0	0	0	-0.146	0	0	-0.075	0.226	0	0.044	-0.031	0
BC-est.	0	0	0	-0.174	0	0	-0.074	0.222	0	0.073	-0.003	0
t-stat	0	0	0	-1.437	0	0	-2.677	11.500*	0	2.618*	-0.145	0
						Post-agLasso: Number of factors=3						
estimate	0	0	0	-0.065	0	0	-0.168	0.222	0	0.080	-0.009	0
t-stat	0	0	0	-0.554	0	0	-6.402*	12.117*	0	2.787*	-0.446	0

Note: BC-est. denotes the bias-corrected estimate. * denotes significance at the 5% level. Lag1, Lag2, and Lag3 refer to the first, second, and third lag of economic growth, respectively.

Table 8: Nonlinear estimation I

	Fert	Popu	Con	Inv	Lag1	Lag2	Lag3	Lag1^2	Gov × Inv
					Number of factors=0				
estimate	-0.145	-0.350	-0.021	0.128	0.128	-0.004	0.031	0.009	-0.002
t-stat	-0.741	-0.729	-2.223*	2.067*	1.476	-0.045	1.306	3.103*	-0.998
					Number of factors=3				
estimate	-0.101	0.052	-0.094	0.180	0.145	-0.047	0.014	0.006	-0.005
t-stat	-0.352	0.122	-5.525*	2.431*	2.070*	-0.769	0.663	3.009*	-2.189*
					Number of factors=5				
estimate	-0.223	0.037	-0.027	0.350	0.082	-0.136	0.015	0.006	-0.003
t-stat	-0.572	0.069	-1.371	3.513*	0.937	-1.687	0.515	2.629*	-1.304
					Number of factors=8				
estimate	0.187	-0.151	-0.029	0.362	0.057	-0.251	0.036	0.005	-0.006
t-stat	0.332	-0.211	-0.996	2.745*	0.520	-2.638*	0.983	1.292	-1.401
					AgLasso: Number of factors=3				
estimate	0.579	-0.318	-0.038	0.303	0.015	-0.087	-0.009	0.003	-0.006
BC-est.	0.512	-0.351	-0.035	0.300	0.045	-0.052	-0.016	0.003	-0.006
t-stat	3.314*	-2.317*	-3.861*	12.84*	1.690	-2.245*	-0.793	1.702	-4.807*
					Post-agLasso: Number of factors=3				
estimate	-0.151	-0.286	-0.104	0.219	0.052	-0.012	-0.003	0.003	-0.008
t-stat	-0.709	-0.839	-5.434*	7.423*	1.703	-0.499	-0.115	1.224	-4.536*

Note: BC-est. denotes the bias-corrected estimate. * denotes significance at the 5% level. Lag1, Lag2, and Lag3 refer to the first, second, and third lag of economic growth, respectively.

Table 9: Nonlinear estimation II

	Life	Con	Gov	Inv	Lag1	Fert^2	Fert × Young	Fert × Life	Fert × Popu	Fert × Lag1	Fert × Lag2
						Number of factors=0					
estimate	-0.247	0.155	0.353	0.473	0.594	-0.489	0.065	0.017	0.564	0.005	0.004
t-stat	-1.540	1.323	1.053	1.921	1.504	-1.871	1.778	0.852	2.273*	0.113	0.094
						Number of factors=3					
estimate	0.179	0.026	0.787	-0.020	0.391	-0.364	0.058	-0.020	0.143	0.010	-0.024
t-stat	0.629	0.137	1.155	-0.056	0.831	-0.974	1.147	-0.607	0.313	0.169	-0.480
						Number of factors=5					
estimate	0.216	-0.046	0.728	0.006	0.244	-0.426	0.057	-0.030	0.082	-0.011	-0.083
t-stat	0.919	-0.247	1.226	0.020	0.606	-1.143	1.139	-0.930	0.231	-0.192	-1.614
						Number of factors=8					
estimate	0.169	-0.107	0.445	0.043	0.065	-0.602	0.085	-0.028	0.231	-0.038	-0.052
t-stat	0.576	-0.504	0.669	0.122	0.150	-1.264	1.277	-0.729	0.571	-0.647	-0.931
						AgLasso: Number of factors=3					
estimate	0.018	-0.035	-0.059	0.084	0.145	0.003	0.003	-0.001	-0.045	-0.014	-0.001
BC-est.	0.009	-0.031	-0.054	0.081	0.178	0.001	0.003	-0.000	-0.049	-0.016	0.003
t-stat	0.566	-2.912*	-1.772	4.472*	3.101*	0.021	0.598	-0.074	-1.694	-1.406	0.632
						Post-agLasso: Number of factors=3					
estimate	-0.043	-0.082	-0.252	0.139	0.321	0.003	0.002	-0.006	-0.033	-0.049	0.004
t-stat	-1.126	-4.875*	-6.435*	6.169*	5.611*	0.066	0.378	-0.865	-1.341	-4.250*	0.936

Note: BC-est. denotes the bias-corrected estimate. * denotes significance at the 5% level. Lag1 and Lag2 refer to the first and second lag of economic growth, respectively.

Supplementary Material for
"Shrinkage Estimation of Dynamic Panel Data Models with Interactive Fixed Effects"

Xun Lu[a], Liangjun Su[b]

[a] *Department of Economics, HKUST*

[b] *School of Economics, Singapore Management University*

This supplementary material provides proofs for the technical lemmas and Corollary 3.4 in the above paper. We also present some primitive conditions to verify some high level conditions in Assumptions A1, A2, A4, and A5 in the text.

C Proofs of the technical lemmas in Appendix A

Proof of Lemma A.1. (i) From the principal component analysis, we have the identity $(NT)^{-1}\hat{\mathbf{Y}}'\hat{\mathbf{Y}}\tilde{F}=\tilde{F}V_{NT}$. Pre-multiplying both sides by $T^{-1}\tilde{F}'$ and using the normalization $T^{-1}\tilde{F}'\tilde{F}=I_R$ yield $T^{-1}\tilde{F}'(NT)^{-1}\hat{\mathbf{Y}}'\hat{\mathbf{Y}}\tilde{F}=V_{NT}$. (i) follows provided $V=\text{plim}V_{NT}$, which we show below.

(ii) Let $\mathbf{e}\equiv\boldsymbol{\varepsilon}+\sum_{k=1}^{K}(\beta_k^0-\tilde{\beta}_k^c)\mathbf{X}_k$. Then $\|\mathbf{e}\|_{\text{sp}}\leq\|\boldsymbol{\varepsilon}\|_{\text{sp}}+\sum_{k=1}^{K}|\beta_k^0-\tilde{\beta}_k^c|\,\|\mathbf{X}_k\|_{\text{sp}}\leq\|\boldsymbol{\varepsilon}\|_{\text{sp}}+\left\|\beta^0-\tilde{\beta}^c\right\|\{\sum_{k=1}^{K}\|\mathbf{X}_k\|_{\text{sp}}^2\}^{1/2}=O_P(\sqrt{N}+\sqrt{T})+O_P(K)=O_P(\sqrt{N}+\sqrt{T})$ by Assumptions A.1(i), (iv) and (v) and A.3(i). Noting that $\hat{\mathbf{Y}}=\mathbf{Y}-\sum_{k=1}^{K}\tilde{\beta}_k^c\mathbf{X}_k=\lambda^0F^{0\prime}+\mathbf{e}$, (i) implies that

$$(T^{-1}\tilde{F}'F^0)\left(N^{-1}\lambda^{0\prime}\lambda^0\right)(T^{-1}F^{0\prime}\tilde{F})+d_{NT}=V_{NT}\xrightarrow{P}V \tag{C.1}$$

where

$$d_{NT}=N^{-1}T^{-2}\tilde{F}'e'e\tilde{F}+(T^{-1}\tilde{F}'F^0)(N^{-1}T^{-1}\lambda^{0\prime}e\tilde{F})+(N^{-1}T^{-1}\tilde{F}'e'\lambda^0)(T^{-1}F^{0\prime}\tilde{F}).$$

Noting that $N^{-1}T^{-2}||\tilde{F}'e'e\tilde{F}||\leq RN^{-1}T^{-1}\{T^{-1}||\tilde{F}||^2\}\|e\|_{\text{sp}}^2=N^{-1}T^{-1}O_P(N+T)=O_P(N^{-1}+T^{-1})$, $N^{-1}T^{-1}||\lambda^{0\prime}e\tilde{F}||\leq N^{-1/2}\{T^{-1/2}||\tilde{F}||\}\{N^{-1/2}T^{-1/2}\left\|\lambda^{0\prime}e\right\|\}=O_P(N^{-1/2})$, and $T^{-1}||F^{0\prime}\tilde{F}||=O_P(1)$, we have $\|d_{NT}\|=O_P(T^{-1}+N^{-1/2})=o_P(1)$. It follows that $(\tilde{F}'F^0/T)\left(\lambda^{0\prime}\lambda^0/N\right)(F^{0\prime}\tilde{F}/T)\xrightarrow{P}V$.

We are left to show that V is the probability limit of V_{NT}. We discuss two cases: $R=R_0$ and $R>R_0$. The first case is studied in Bai (2003, Lemma A.3) who characterizes V as a diagonal matrix consisting of the R_0 eigenvalues of $\Sigma_{\lambda^0}\Sigma_{F^0}$, arranged in descending order. In the second case, observe that $(T^{-1}\tilde{F}'F^0)\left(N^{-1}\lambda^{0\prime}\lambda^0\right)(T^{-1}F^{0\prime}\tilde{F})$ has rank R_0 at most in both finite and large samples. Let $\Delta_{lNT}=T^{-1}F^{0\prime}\tilde{F}_{(l)}$ for $l=1,2$, and $\hat{\Sigma}_\lambda=N^{-1}\lambda^{0\prime}\lambda^0$. Then

$$(T^{-1}\tilde{F}'F^0)\left(N^{-1}\lambda^{0\prime}\lambda^0\right)(T^{-1}F^{0\prime}\tilde{F})=\left[\begin{array}{cc}\Delta_{1NT}'\hat{\Sigma}_\lambda\Delta_{1NT} & \Delta_{1NT}'\hat{\Sigma}_\lambda\Delta_{2NT}\\ \Delta_{2NT}'\hat{\Sigma}_\lambda\Delta_{1NT} & \Delta_{2NT}'\hat{\Sigma}_\lambda\Delta_{2NT}\end{array}\right].$$

By Lemma A.3(ii) of Bai (2003), $\Delta'_{1NT}\hat{\Sigma}_\lambda\Delta_{1NT} \xrightarrow{P} V_{11}$, which has full rank R_0 under Assumptions A.1(ii) and (iii). This ensures that $(\tilde{F}'F^0/T)\left(\lambda^{0\prime}\lambda^0/N\right)(F^{0\prime}\tilde{F}/T)$ has rank R_0 in large samples and $\Delta'_{2NT}\hat{\Sigma}_\lambda\Delta_{2NT} \xrightarrow{P} 0$. Then $\Delta'_{1NT}\hat{\Sigma}_\lambda\Delta_{2NT} \xrightarrow{P} 0$ by Cauchy-Schwarz (CS hereafter) inequality. It follows that $V = \begin{bmatrix} V_{11} & 0 \\ 0 & 0 \end{bmatrix}$.

(iii) From the above proof and the fact that $\hat{\Sigma}_\lambda$ is asymptotically nonsingular by Assumption A.1(iii), we have $\Delta_{2NT} \xrightarrow{P} 0$ and $\Delta_{1NT} \xrightarrow{P} \Delta_1$ where Δ_1 is an $R_0 \times R_0$ full rank matrix.

(iv) This follows from (iii), Assumption A.1(iii), and Slutsky lemma. ■

Proof of Lemma A.2. (i) Using $\mathbf{Y} - \sum_{k=1}^{K}\tilde{\beta}_k^c\mathbf{X}_k = \lambda^0F^{0\prime} + \boldsymbol{\varepsilon} + \sum_{k=1}^{K}(\beta_k^0 - \tilde{\beta}_k^c)\mathbf{X}_k$ and $F^0H = (NT)^{-1}F^0\lambda^{0\prime}\lambda^0F^{0\prime}\tilde{F}$ yields

$$\begin{aligned}
\hat{F} - F^0H &= (NT)^{-1}\left(\mathbf{Y} - \sum_{k=1}^{K}\tilde{\beta}_k^c\mathbf{X}_k\right)'\left(\mathbf{Y} - \sum_{k=1}^{K}\tilde{\beta}_k^c\mathbf{X}_k\right)\tilde{F} - F^0H \\
&= (NT)^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\tilde{F} + (NT)^{-1}F^0\lambda^{0\prime}\boldsymbol{\varepsilon}\tilde{F} + (NT)^{-1}\boldsymbol{\varepsilon}'\lambda^0F^{0\prime}\tilde{F} \\
&\quad + (NT)^{-1}\sum_{k=1}^{K}\left(\beta_k^0 - \tilde{\beta}_k^c\right)\mathbf{X}'_k\sum_{l=1}^{K}\left(\beta_l^0 - \tilde{\beta}_l^c\right)\mathbf{X}_l\tilde{F} \\
&\quad + (NT)^{-1}F^0\lambda^{0\prime}\sum_{k=1}^{K}\left(\beta_k^0 - \tilde{\beta}_k^c\right)\mathbf{X}_k\tilde{F} + (NT)^{-1}\sum_{k=1}^{K}\left(\beta_k^0 - \tilde{\beta}_k^c\right)\mathbf{X}'_k\lambda^0F^{0\prime}\tilde{F} \\
&\quad + (NT)^{-1}\boldsymbol{\varepsilon}'\sum_{k=1}^{K}\left(\beta_k^0 - \tilde{\beta}_k^c\right)\mathbf{X}_k\tilde{F} + (NT)^{-1}\sum_{k=1}^{K}\left(\beta_k^0 - \tilde{\beta}_k^c\right)\mathbf{X}'_k\boldsymbol{\varepsilon}\tilde{F} \\
&\equiv a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 + a_8, \text{ say.} \qquad \text{(C.2)}
\end{aligned}$$

It follows that $T^{-1}||\hat{F} - F^0H||^2 \leq 8T^{-1}\sum_{l=1}^{8}\|a_l\|^2$ by CS inequality. Similarly, we can write $\hat{F}_{(1)} - F^0H_{(1)} = \sum_{l=1}^{8}a_{l(1)}$, where $a_l = (a_{l(1)}, a_{l(2)})$ and $a_{l(1)}$ denotes a $T \times R_0$ submatrix of a_l for $l = 1, 2, ..., 8$.

By Bai and Ng (2002, pp.213-214), $T^{-1}\|a_l\|^2 = O_P\left(\delta_{NT}^{-2}\right)$ for $l = 1, 2, 3$ under Assumptions A.1(ii)-(iii) and A.2(i)-(iii) as our assumptions also ensure their Lemma 1 to hold. By the facts that $||\beta^0 - \tilde{\beta}^c|| = O_P((NT/K)^{-1/2})$, $T^{-1}||\tilde{F}||^2 = O_P(1)$, $(NT)^{-1/2}\|\mathbf{X}_k\| = O_P(1)$, $N^{-1}\left\|\lambda^0\right\|^2 = O_P(1)$, $T^{-1}||\tilde{F}'F^0|| = O_P(1)$, and that $\sum_{k=1}^{K}\|\mathbf{X}'_k\boldsymbol{\varepsilon}\|_{\text{sp}}^2 = O_P(NTK(N+T))$ under Assumption A.1, we have

$$\begin{aligned}
T^{-1}\|a_4\|^2 &\leq K^2\left\|\beta^0 - \tilde{\beta}^c\right\|^4\left\{T^{-1}\left\|\tilde{F}\right\|^2\right\}\left\{(NTK)^{-1}\sum_{k=1}^{K}\|\mathbf{X}_k\|^2\right\}^2 = O_P(K^4(NT)^{-2}), \\
T^{-1}\|a_5\|^2 &\leq K\left\|\beta^0 - \tilde{\beta}^c\right\|^2\left\{T^{-1}\left\|\tilde{F}\right\|^2\right\}\left\{(NTK)^{-1}\sum_{k=1}^{K}\|\mathbf{X}_k\|^2\right\}(NT)^{-1}\left\|F^0\lambda^{0\prime}\right\|^2 = O_P(K^2(NT)^{-1}), \\
T^{-1}\|a_6\|^2 &\leq K\left\|\beta^0 - \tilde{\beta}^c\right\|^2\left\{T^{-1}\left\|\tilde{F}'F^0\right\|\right\}^2\left\{N^{-1}\left\|\lambda^0\right\|^2\right\}\left\{(NTK)^{-1}\sum_{k=1}^{K}\|\mathbf{X}_k\|^2\right\} = O_P(K^2(NT)^{-1}), \\
T^{-1}\|a_7\|^2 &\leq N^{-2}T^{-2}\eta_{NT}\left\|\beta^0 - \tilde{\beta}^c\right\|^2\left\{T^{-1}\left\|\tilde{F}\right\|^2\right\}\eta_{NT}^{-1}\sum_{k=1}^{K}\|\mathbf{X}'_k\boldsymbol{\varepsilon}\|^2 = O_P(K^2(NT)^{-2}(N+T)),
\end{aligned}$$

and

$$T^{-1}\left\|a_8\right\|^2 \leq N^{-2}T^{-2}\eta_{NT}\left\|\beta^0-\tilde{\beta}^c\right\|^2\left\{T^{-1}\left\|\tilde{F}\right\|^2\right\}\eta_{NT}^{-1}\sum_{k=1}^{K}\left\|\mathbf{X}_k'\boldsymbol{\varepsilon}\right\|^2 = O_P(K^2\left(NT\right)^{-2}\left(N+T\right)),$$

where $\eta_{NT} = NTK(N+T)$. Hence $T^{-1}||\hat{F}-F^0H||^2 = O_P(\delta_{NT}^{-2}+K^2\left(NT\right)^{-1}) = O_P(\delta_{NT}^{-2})$ under Assumption A.3(i).

(ii) By the definition $\hat{F} = \left(NT\right)^{-1}\hat{\mathbf{Y}}'\hat{\mathbf{Y}}\tilde{F}$ and the identity $\left(NT\right)^{-1}\hat{\mathbf{Y}}'\hat{\mathbf{Y}}\tilde{F} = \tilde{F}V_{NT}$, $T^{-1}||\hat{F}-F^0H||^2 = T^{-1}||\tilde{F}V_{NT}-F^0H||^2 = T^{-1}||\tilde{F}_{(1)}V_{NT,11}-F^0H_{(1)}||^2 + T^{-1}||\tilde{F}_{(2)}V_{NT,22}-F^0H_{(2)}||^2$. Then (i) implies that $T^{-1}||\tilde{F}_{(1)}V_{NT,11}-F^0H_{(1)}||^2 = O_P(\delta_{NT}^{-2})$ and $T^{-1}||\tilde{F}_{(2)}V_{NT,22}-F^0H_{(2)}||^2 = O_P(\delta_{NT}^{-2})$. Noting that $V_{NT,11}$ is asymptotically nonsingular by Lemma A.1(i), we have $T^{-1}||\tilde{F}_{(1)}-F^0H_{(1)}V_{NT,11}^{-1}||^2 = O_P(\delta_{NT}^{-2})$. In addition,

$$\begin{aligned}T^{-1}\left\|F^0H_{(2)}\right\|^2 &\leq 2T^{-1}\left\|\tilde{F}_{(2)}V_{NT,22}-F^0H_{(2)}\right\|^2+2T^{-1}\left\|\tilde{F}_{(2)}V_{NT,22}\right\|^2\\ &= O_P\left(\delta_{NT}^{-2}\right)+O_P\left(T^{-2}+N^{-1}\right)=O_P\left(\delta_{NT}^{-2}\right)\end{aligned}$$

because $T^{-1}\left\|\tilde{F}_{(2)}V_{NT,22}\right\|^2 \leq \left[\mu_{\max}\left(V_{NT,22}\right)\right]^2T^{-1}\left\|\tilde{F}_{(2)}\right\|^2 = \left(R-R_0\right)\left[\mu_{\max}\left(V_{NT,22}\right)\right]^2$ and

$$\begin{aligned}\mu_{\max}\left(V_{NT,22}\right) &= \mu_{R_0+1}\left(V_{NT}\right)\leq\mu_{R_0+1}\left(\left(T^{-1}\tilde{F}'F^0\right)\left(N^{-1}\lambda^{0\prime}\lambda^0\right)\left(T^{-1}F^{0\prime}\tilde{F}\right)\right)+\left\|d_{NT}\right\|\\ &= 0+O_P(T^{-1}+N^{-1/2})\end{aligned}$$

by (C.1), the calculation below it, Weyl's inequality and the fact that $\left\|A\right\|_{\mathrm{sp}}\leq\left\|A\right\|$. Observing that $T^{-1}\left\|F^0H_{(2)}\right\|^2 = T^{-1}\mathrm{tr}(H_{(2)}H_{(2)}'F^{0\prime}F^0)\geq\mu_{\min}\left(T^{-1}F^{0\prime}F^0\right)||H_{(2)}||^2$, it follows that $||H_{(2)}||\leq O_P\left(\delta_{NT}^{-1}\right)/[\mu_{\min}\left(T^{-1}F^{0\prime}F^0\right)]^{1/2}=O_P\left(\delta_{NT}^{-1}\right)$.

(iii) Writing $T^{-1}(\hat{F}-F^0H)'F^0=\sum_{l=1}^{8}T^{-1}a_l'F^0=\sum_{l=1}^{8}A_l$, it suffices to show that $\left\|A_l\right\|=O_P(\delta_{NT}^{-2}+(NT/K)^{-1/2})$ for $l=1,3,...,8$, $\left\|A_2\right\|=O_P(N^{-1/2})$, and $||A_{2(1)}||=O_P(\delta_{NT}^{-2})$, where $A_{2(1)}$ is defined analogously as A_2 with a_2 being replaced by $a_{2(1)}$. By the definitions of a_l's in (C.2) and Assumption A.1, we can readily show that $\left\|A_1\right\|=O_P(\delta_{NT}^{-2})$, $\left\|A_2\right\|=O_P(N^{-1/2})$, $\left\|A_3\right\|=O_P((NT)^{-1/2})$, $\left\|A_4\right\|=O_P(K^2\left(NT\right)^{-1})$, and $\left\|A_l\right\|=O_P(K\left(NT\right)^{-1})$ for $l=7,8$. For A_5, we have

$$\begin{aligned}\left\|A_5\right\| &= N^{-1}T^{-2}\left\|F^{0\prime}F^0\lambda^{0\prime}\sum_{k=1}^{K}\left(\beta_k^0-\tilde{\beta}_k^c\right)\mathbf{X}_k\tilde{F}\right\|=N^{-1}T^{-2}\left\|F^{0\prime}F^0\sum_{j=1}^{N}\lambda_j^0\left(\beta^0-\tilde{\beta}^c\right)'X_j'\tilde{F}\right\|\\ &\leq N^{-1}T^{-1/2}\left\{T^{-1}\left\|F^{0\prime}F^0\right\|T^{-1/2}\left\|\tilde{F}\right\|\right\}\upsilon_{1NT},\end{aligned}$$

where $\upsilon_{1NT}=\left\|\sum_{j=1}^{N}\lambda_j^0\left(\beta^0-\tilde{\beta}^c\right)'X_j'\right\|$. Note that

$$\begin{aligned}\upsilon_{1NT}^2 &= \left(\beta^0-\tilde{\beta}^c\right)'\sum_{i=1}^{N}\sum_{j=1}^{N}X_i'X_j\lambda_j^{0\prime}\lambda_i^0\left(\beta^0-\tilde{\beta}^c\right)\leq\left\|\beta^0-\tilde{\beta}^c\right\|^2\mu_{\max}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}X_i'X_j\lambda_j^{0\prime}\lambda_i^0\right)\\ &= O_P\left(\left(NT/K\right)^{-1}\right)O_P\left(N^2T\right)=O_P\left(NK\right),\end{aligned}$$

where we use the fact that under Assumptions A.1(iii) and (viii)

$$\begin{aligned}
\mu_{\max}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}X_j'X_i\lambda_j^{0\prime}\lambda_i^0\right) &= \max_{\varkappa=(\varkappa_1,\ldots,\varkappa_K)':\|\varkappa\|=1}\varkappa'\sum_{i=1}^{N}\sum_{j=1}^{N}X_j'X_i\lambda_j^{0\prime}\lambda_i^0\varkappa \\
&= \max_{\varkappa=(\varkappa_1,\ldots,\varkappa_K)':\|\varkappa\|=1}\operatorname{tr}\left(\sum_{m=1}^{K}\sum_{n=1}^{K}\varkappa_m\varkappa_n X_{(m)}X_{(n)}'\lambda^0\lambda^{0\prime}\right) \\
&\le \left\|\lambda^0\right\|^2\max_{\varkappa=(\varkappa_1,\ldots,\varkappa_K)':\|\varkappa\|=1}\sum_{m=1}^{K}\sum_{n=1}^{K}\varkappa_m\varkappa_n\operatorname{tr}\left(X_{(m)}X_{(n)}'\right) \\
&= \left\|\lambda^0\right\|^2\max_{\varkappa=(\varkappa_1,\ldots,\varkappa_K)':\|\varkappa\|=1}\sum_{m=1}^{K}\sum_{n=1}^{K}\sum_{i=1}^{N}\sum_{t=1}^{T}\varkappa_m\varkappa_n X_{it,m}X_{it,n} \\
&= \left\|\lambda^0\right\|^2\mu_{\max}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}X_{it}'\right) = O_P(N)\,O_P(NT).
\end{aligned}$$

It follows that $\|A_5\| = N^{-1}T^{-1/2}O_P(1)\,O_P\left(N^{1/2}K^{1/2}\right) = O_P\left((NT/K)^{-1/2}\right)$. Similarly, we can show that $\|A_6\| = O_P\left((NT/K)^{-1/2}\right)$. It follows that

$$T^{-1}(\hat{F}-F^0H)'F^0 = O_P(N^{-1/2}+(NT/K)^{-1/2}) = O_P(N^{-1/2}) \text{ under Assumption A.3(i),}$$

implying the second part of (iii). In addition, by (ii) we have

$$\begin{aligned}
\left\|A_{2(1)}\right\| &= N^{-1}T^{-2}\left\|\tilde{F}_{(1)}'\varepsilon'\lambda^0F^{0\prime}F^0\right\| \\
&\le N^{-1}T^{-2}\left\{\left\|\left(\tilde{F}_{(1)}-F^0H_{(1)}V_{NT,11}^{-1}\right)\varepsilon'\lambda^0F^{0\prime}F^0\right\| + \left\|V_{NT,11}^{-1}H_{(1)}'F^0\varepsilon'\lambda^0F^{0\prime}F^0\right\|\right\} \\
&\le N^{-1}T^{-2}\left\{\left\|\tilde{F}_{(1)}-F^0H_{(1)}V_{NT,11}^{-1}\right\|\left\|\varepsilon'\lambda^0\right\|\left\|F^{0\prime}F^0\right\| + \left\|V_{NT,11}^{-1}H_{(1)}\right\|\left\|F^0\varepsilon'\lambda^0\right\|\left\|F^{0\prime}F^0\right\|\right\} \\
&= N^{-1}T^{-2}\left\{O_P(T^{1/2}\delta_{NT}^{-1})O_P(N^{1/2}T^{1/2})O_P(T) + O_P(1)\,O_P(N^{1/2}T^{1/2})O_P(T)\right\} = O_P\left(\delta_{NT}^{-2}\right).
\end{aligned}$$

It follows that $T^{-1}(\hat{F}_{(1)}-F^0H_{(1)})'F^0 = O_P(\delta_{NT}^{-2}+(NT/K)^{-1/2})$.

(iv) Noting that $T^{-1}(\hat{F}-F^0H)'\hat{F} = T^{-1}(\hat{F}-F^0H)'(\hat{F}-F^0H) + T^{-1}(\hat{F}-F^0H)'F^0H$, the first two parts of the result follow from (i) and (iii). To prove the third part, note that $T^{-1}(\hat{F}-F^0H)'\hat{F}_{(2)} = (T^{-1}(\hat{F}_{(1)}-F^0H_{(1)})'\hat{F}_{(2)},\ T^{-1}(\hat{F}_{(2)}-F^0H_{(2)})'\hat{F}_{(2)})$. By the triangle inequality, the submultiplicative property of the Frobenius norm, and (i)-(iii)

$$\begin{aligned}
T^{-1}\left\|(\hat{F}_{(1)}-F^0H_{(1)})'\hat{F}_{(2)}\right\| &\le T^{-1}\left\|(\hat{F}_{(1)}-F^0H_{(1)})'(\hat{F}_{(2)}-F^0H_{(2)})\right\| \\
&\quad + T^{-1}\left\|(\hat{F}_{(1)}-F^0H_{(1)})'F^0\right\|\left\|H_{(2)}\right\| \\
&= O_P(\delta_{NT}^{-2}) + O_P(\delta_{NT}^{-2}+(NT/K)^{-1/2})O_P(\delta_{NT}^{-1}) = O_P(\delta_{NT}^{-2}),
\end{aligned}$$

and

$$\begin{aligned}
T^{-1}\left\|(\hat{F}_{(2)}-F^0H_{(2)})'\hat{F}_{(2)}\right\| &\le T^{-1}\left\|(\hat{F}_{(2)}-F^0H_{(2)})'(\hat{F}_{(2)}-F^0H_{(2)})\right\| \\
&\quad + T^{-1}\left\|(\hat{F}_{(2)}-F^0H_{(2)})'F^0\right\|\left\|H_{(2)}\right\| \\
&= O_P(\delta_{NT}^{-2}) + O_P(N^{-1/2})O_P(\delta_{NT}^{-1}) = O_P(\delta_{NT}^{-2}).
\end{aligned}$$

It follows that $T^{-1}(\hat{F}-F^0H)'\hat{F}_{(2)}=O_P(\delta_{NT}^{-2})$. This proves (iv).

(v) In view of the fact that $T^{-1}(\hat{F}'\hat{F}-H'F^{0\prime}F^0H)=T^{-1}(\hat{F}-F^0H)'(\hat{F}-F^0H)+T^{-1}(\hat{F}-F^0H)'F^0H$ $+T^{-1}\left(F^0H\right)'(\hat{F}-F^0H)$, the result follows from (i) and (iii).

(vi) Observe that

$$\begin{aligned} P_{\hat{F}_{(1)}}-P_{F^*_{(1)}} &= \hat{F}_{(1)}\left(\hat{F}'_{(1)}\hat{F}_{(1)}\right)^{-1}\hat{F}'_{(1)}-F^*_{(1)}\left(F^{*\prime}_{(1)}F^*_{(1)}\right)^{-1}F^{*\prime}_{(1)} \\ &= \left(\hat{F}_{(1)}-F^*_{(1)}\right)\left(\hat{F}'_{(1)}\hat{F}_{(1)}\right)^{-1}\left(\hat{F}_{(1)}-F^*_{(1)}\right)'+\left(\hat{F}_{(1)}-F^*_{(1)}\right)\left(\hat{F}'_{(1)}\hat{F}_{(1)}\right)^{-1}F^{*\prime}_{(1)} \\ &\quad +F^*_{(1)}\left(\hat{F}'_{(1)}\hat{F}_{(1)}\right)^{-1}\left(\hat{F}_{(1)}-F^*_{(1)}\right)'+F^*_{(1)}\left[\left(\hat{F}'_{(1)}\hat{F}_{(1)}\right)^{-1}-\left(F^{*\prime}_{(1)}F^*_{(1)}\right)^{-1}\right]F^{*\prime}_{(1)} \\ &\equiv b_1+b_2+b_3+b_4,\ \text{say.} \end{aligned} \tag{C.3}$$

By (v), $T^{-1}||\hat{F}_{(1)}-F^*_{(1)}||^2=O_P(\delta_{NT}^{-2})$ and $T^{-1}||\hat{F}'_{(1)}\hat{F}_{(1)}-F^{*\prime}_{(1)}F^*_{(1)}||=O_P(\delta_{NT}^{-2}+(NT/K)^{-1/2})$. With these, one can readily show that $\|b_1\|=O_P(\delta_{NT}^{-2})$ and $\|b_4\|=O_P(\delta_{NT}^{-2}+(NT/K)^{-1/2})$. For b_2, we have $\|b_2\|\leq T^{-1/2}||\hat{F}_{(1)}-F^*_{(1)}||\ ||(T^{-1}\hat{F}'_{(1)}\hat{F}_{(1)})^{-1}||T^{-1/2}||F^*_{(1)}||=O_P(\delta_{NT}^{-1})$. Noticing that $b_3=b_2'$, we have completed the proof of (vi).

(vii) Note that $T^{-1}H'F^{0\prime}F^0H=(\tilde{F}'F^0/T)\left(\lambda^{0\prime}\lambda^0/N\right)\left(F^{0\prime}F^0/T\right)\left(\lambda^{0\prime}\lambda^0/N\right)(F^{0\prime}\tilde{F}/T)$ has rank R_0 in large samples by Lemma A.1(iii). As a result, $\breve{\tau}_l$ converges in probability to some positive number τ_l^* for $l=1,...,R_0$ and $\breve{\tau}_l=0$ for $l=R_0+1,...,R$. In addition, by (iv) and perturbation theory for eigenvalue problems (e.g., Stewart and Sun (1990, p.203)),

$$|\tau_l-\breve{\tau}_l|\leq\left\|T^{-1}\hat{F}'\hat{F}-T^{-1}H'F^{0\prime}F^0H\right\|=O_P(N^{-1/2})\text{ for }l=1,...,R_0$$

and

$$|\tau_l|=|\tau_l-\breve{\tau}_l|\leq\left\|T^{-1}\hat{F}'\hat{F}-T^{-1}H'F^{0\prime}F^0H\right\|=O_P(N^{-1/2})\text{ for }l=R_0+1,...,R.$$

It follows that $\tau_l=\tau_l^*+o_P(1)$ for $l=1,...,R_0$ and $\tau_l=O_P(N^{-1/2})$ for $l=R_0+1,...,R$. ■

Proof of Lemma A.3. (i) By (C.3) we have

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^N X'_{i(1)}\left(P_{\hat{F}_{(1)}}-P_{F^0_{(1)}}\right)\left(\hat{F}_{(1)}-F^*_{(1)}\right)\lambda^*_{i(1)}=\sum_{l=1}^4\frac{1}{\sqrt{NT}}\sum_{i=1}^N X'_{i(1)}b_l\left(\hat{F}_{(1)}-F^*_{(1)}\right)\lambda^*_{i(1)}\equiv\sum_{l=1}^4 B_{lNT}.$$

Let c_{K_0} be an arbitrary $K_0\times 1$ nonrandom vector with $\|c_{K_0}\|=1$. For B_{1NT}, noting that $|\mathrm{tr}\,(A)|\leq\mathrm{rank}(A)\,\|A\|$ and $\|b_1\|=O_P\left(\delta_{NT}^{-2}\right)$, we can apply Lemma A.2(i), and Assumptions A.1(iii) and (viii) and A.3(i) to obtain

$$\begin{aligned}
\left|c_{K_0}' B_{1NT}\right| &= (NT)^{-1/2} \left| \operatorname{tr}\left(\sum_{i=1}^{N} \lambda_{i(1)}^{*} c_{K_0}' X_{i(1)}' b_l \left(\hat{F}_{(1)} - F_{(1)}^{*} \right) \right) \right| \\
&\leq R_0 (NT)^{-1/2} \left\| \sum_{i=1}^{N} \lambda_{i(1)}^{*} c_{K_0}' X_{i(1)}' b_l \left(\hat{F}_{(1)} - F_{(1)}^{*} \right) \right\| \\
&\leq R_0 (NT)^{-1/2} \left\| b_1 \right\| \left\| \hat{F}_{(1)} - F_{(1)}^{*} \right\| \left\| \sum_{i=1}^{N} \lambda_{i(1)}^{*} c_{K_0}' X_{i(1)}' \right\| \\
&= (NT)^{-1/2} O_P\left(\delta_{NT}^{-2}\right) O_P\left(T^{1/2}\delta_{NT}^{-1}\right) O_P(NT^{1/2}) = O_P\left(N^{1/2}T^{1/2}\delta_{NT}^{-3}\right) = o_P(1),
\end{aligned}$$

where we use the fact that $\left\|\sum_{i=1}^{N} \lambda_{i(1)}^{*} c_{K_0}' X_{i(1)}'\right\| = O_P\left(NT^{1/2}\right)$ by analogous arguments as used in the study of ς_{2NT} in the proof of Proposition B.1. It follows that $\|B_{1NT}\| = o_P(1)$. Similarly, using that $\|b_4\| = O_P(\delta_{NT}^{-2} + (NT/K)^{-1/2})$, Lemmas A.2(i), (iii) and (iv), and Assumptions A.1(iii) and (viii) and A.3(i), we have

$$\begin{aligned}
\left|c_{K_0}' B_{4NT}\right| &\leq R_0 (NT)^{-1/2} \left\| b_4 \right\| \left\| \hat{F}_{(1)} - F_{(1)}^{*} \right\| \left\| \sum_{i=1}^{N} \lambda_{i(1)}^{*} c_{K_0}' X_{i(1)}' \right\| \\
&= (NT)^{-1/2} O_P\left(\delta_{NT}^{-2} + (NT/K)^{-1/2}\right) O_P\left(T^{1/2}\delta_{NT}^{-1}\right) O_P\left(NT^{1/2}\right) \\
&= O_P\left(N^{1/2}T^{1/2}\delta_{NT}^{-3} + K^{1/2}\delta_{NT}^{-1}\right) = o_P(1),
\end{aligned}$$

and

$$\begin{aligned}
\left|c_{K_0}' B_{2NT}\right| &= \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} c_{K_0}' X_{i(1)}' \left(\hat{F}_{(1)} - F_{(1)}^{*} \right) \left(\hat{F}_{(1)}' \hat{F}_{(1)} \right)^{-1} F_{(1)}^{*\prime} \left(\hat{F}_{(1)} - F_{(1)}^{*} \right) \lambda_{i(1)}^{*} \right\| \\
&\leq R_0 (NT)^{-1/2} \left\| \hat{F}_{(1)} - F_{(1)}^{*} \right\| T^{-1} \left\| F_{(1)}^{*\prime} \left(\hat{F}_{(1)} - F_{(1)}^{*} \right) \right\| \left\| \left(T^{-1} \hat{F}_{(1)}' \hat{F}_{(1)} \right)^{-1} \right\| \left\| \sum_{i=1}^{N} \lambda_{i(1)}^{*} c_{K_0}' X_{i(1)}' \right\| \\
&= (NT)^{-1/2} O_P\left(T^{1/2}\delta_{NT}^{-1}\right) O_P\left(\delta_{NT}^{-2} + (NT/K)^{-1/2}\right) O_P(1)\, O_P\left(NT^{1/2}\right) \\
&= O_P\left(N^{1/2}T^{1/2}\delta_{NT}^{-3} + K^{1/2}\delta_{NT}^{-1}\right) = o_P(1).
\end{aligned}$$

For B_{3NT}, using Lemmas A.2(i) and (v), we can readily show that $B_{3NT} = \bar{B}_{3NT} + \mathbf{O}_P(N^{1/2}T^{1/2}\delta_{NT}^{-2}(\delta_{NT}^{-2} + (NT/K)^{-1/2})) = \bar{B}_{3NT} + \mathbf{o}_P(1)$, where

$$\bar{B}_{3NT} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_{i(1)}' F_{(1)}^{*} \left(F_{(1)}^{*\prime} F_{(1)}^{*} \right)^{-1} \left(\hat{F}_{(1)} - F_{(1)}^{*} \right)' \left(\hat{F}_{(1)} - F_{(1)}^{*} \right) \lambda_{i(1)}^{*}.$$

By (C.2), we can write $\bar{B}_{3NT}$ as $\bar{B}_{3NT} = \sum_{k=1}^{8}\sum_{l=1}^{8} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_{i(1)}' F_{(1)}^{*} (F_{(1)}^{*\prime} F_{(1)}^{*})^{-1} a_{k(1)}' a_{l(1)} \lambda_{i(1)}^{*} \equiv \sum_{k=1}^{8}\sum_{l=1}^{8} \bar{B}_{3NT}(k,l)$, say. Note that

$$\bar{B}_{3NT}(3,3) = \frac{1}{N^{5/2}T^{5/2}} \sum_{i=1}^{N} X_{i(1)}' F_{(1)}^{*} (F_{(1)}^{*\prime} F_{(1)}^{*})^{-1} \tilde{F}_{(1)}' F^{0} \lambda^{0\prime} \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \lambda^{0} F^{0\prime} \tilde{F}_{(1)} \lambda_{i(1)}^{*} = \mathbb{B}_{1NT}.$$

In addition, one can readily show that $\bar{B}_{3NT}(k,l) = \mathbf{o}_P(1)$ for $k,l = 1,...,8$ with $k \neq l$ or $k = l \neq 3$. For example,

$$
\begin{aligned}
\left\|\bar{B}_{3NT}(1,1)\right\| &= \frac{1}{N^{5/2}T^{5/2}}\left\|\sum_{i=1}^{N} X_{i(1)}' F_{(1)}^{*}\left(F_{(1)}^{*\prime}F_{(1)}^{*}\right)^{-1}\tilde{F}_{(1)}'\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\tilde{F}_{(1)}\lambda_{i(1)}^{*}\right\| \\
&\leq \frac{K_0}{N^{5/2}T^{5/2}}\left\|\tilde{F}_{(1)}'\boldsymbol{\varepsilon}'\right\|^2 \|\boldsymbol{\varepsilon}\|_{\mathrm{sp}}^2 \left\|F_{(1)}^{*}\left(F_{(1)}^{*\prime}F_{(1)}^{*}\right)^{-1}\right\|\left\{\sum_{i=1}^{N}\left\|X_{i(1)}\right\|\left\|\lambda_{i(1)}^{*}\right\|\right\} \\
&= N^{-5/2}T^{-5/2}O_P(NT)\,O_P(N+T)\,O_P\left(T^{-1/2}\right)O_P(K_0^{1/2}NT^{1/2}) \\
&= O_P\left(K_0^{1/2}(N^{-1/2}T^{-1/2} + N^{1/2}T^{-3/2})\right) = o_P(1),
\end{aligned}
$$

where we use the fact that $\left\|\tilde{F}_{(1)}'\boldsymbol{\varepsilon}'\right\| \leq \left\|V_{NT,11}^{-1}H_{(1)}'F^0\boldsymbol{\varepsilon}'\right\| + \left\|\left(\tilde{F}_{(1)} - F^0 H_{(1)}V_{NT,11}^{-1}\right)\boldsymbol{\varepsilon}'\right\| \leq \left\|V_{NT,11}^{-1}\right\| \times \left\|H_{(1)}\right\|\left\|F^0\boldsymbol{\varepsilon}'\right\| + R_0\left\|\tilde{F}_{(1)} - F^0H_{(1)}V_{NT,11}^{-1}\right\|\|\boldsymbol{\varepsilon}\|_{\mathrm{sp}} = O_P(N^{1/2}T^{1/2} + T^{1/2}\delta_{NT}^{-1}\left(N^{1/2}+T^{1/2}\right)) = O_P(N^{1/2}T^{1/2})$ by Lemmas A.1(iii) and A.2(ii) and Assumptions A.1(v) and (vii). Similarly

$$
\begin{aligned}
\left\|\bar{B}_{3NT}(2,2)\right\| &= \frac{1}{N^{5/2}T^{5/2}}\left\|\sum_{i=1}^{N} X_{i(1)}' F_{(1)}^{*}\left(F_{(1)}^{*\prime}F_{(1)}^{*}\right)^{-1}\tilde{F}_{(1)}'\boldsymbol{\varepsilon}'\lambda^0 F^{0\prime}F^0\lambda^{0\prime}\boldsymbol{\varepsilon}\tilde{F}_{(1)}\lambda_{i(1)}^{*}\right\| \\
&\leq \frac{1}{N^{5/2}T^{5/2}}\left\|\tilde{F}_{(1)}'\boldsymbol{\varepsilon}'\lambda^0\right\|^2 \left\|F^0\right\|^2 \left\|F_{(1)}^{*}\left(F_{(1)}^{*\prime}F_{(1)}^{*}\right)^{-1}\right\|\left\{\sum_{i=1}^{N}\left\|X_{i(1)}\right\|\left\|\lambda_{i(1)}^{*}\right\|\right\} \\
&= N^{-5/2}T^{-5/2}O_P\left(NT^2\delta_{NT}^{-2}\right)O_P(T)\,O_P\left(T^{-1/2}\right)O_P(K_0^{1/2}NT^{1/2}) \\
&= O_P\left(K_0^{1/2}N^{-1/2}T^{1/2}\delta_{NT}^{-2}\right) = o_P(1),
\end{aligned}
$$

where we use the fact that $\left\|\tilde{F}_{(1)}'\boldsymbol{\varepsilon}'\lambda^0\right\| \leq \left\|V_{NT,11}^{-1}H_{(1)}'F^0\boldsymbol{\varepsilon}'\lambda^0\right\| + \left\|(\tilde{F}_{(1)} - F^0H_{(1)}V_{NT,11}^{-1})\boldsymbol{\varepsilon}'\lambda^0\right\| \leq \left\|V_{NT,11}^{-1}\right\| \times \left\|H_{(1)}\right\|\left\|F^0\boldsymbol{\varepsilon}'\lambda^0\right\| + \left\|\tilde{F}_{(1)} - F^0H_{(1)}V_{NT,11}^{-1}\right\|\left\|\boldsymbol{\varepsilon}'\lambda^0\right\| = O_P(N^{1/2}T^{1/2} + T^{1/2}\delta_{NT}^{-1}N^{1/2}T^{1/2}) = O_P\left(N^{1/2}T\delta_{NT}^{-1}\right)$ by Lemmas A.1(iii) and A.2(ii) and Assumption A.1(vii). Consequently, $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} X_{i(1)}'(P_{\hat{F}_{(1)}} - P_{F_{(1)}^{*}})(\hat{F}_{(1)} - F_{(1)}^{*})\lambda_{i(1)}^{*} = \mathbb{B}_{1NT} + \mathbf{o}_P(1)$.

(ii) Let $\hat{\Sigma}_{\hat{F}_{(1)}} = T^{-1}\hat{F}_{(1)}'\hat{F}_{(1)}$. Let c_{K_0} be an arbitrary $K_0 \times 1$ nonrandom vector with $\|c_{K_0}\| = 1$. Let $A_{NT} = c_{K_0}'\frac{1}{NT}\sum_{i=1}^{N} X_{i(1)}'(\hat{F}_{(1)} - F_{(1)}^{*})\hat{\Sigma}_{\hat{F}_{(1)}}^{-1}F_{(1)}^{*\prime}\varepsilon_i$. It suffices to show that $A_{NT} = O_P(K_0^{1/2}T^{1/2}\delta_{NT}^{-2})$. Note that

$$
|A_{NT}| = (NT)^{-1}\,\mathrm{tr}\left(\hat{\Sigma}_{\hat{F}_{(1)}}^{-1}H_{(1)}'\sum_{i=1}^{N}F^{0\prime}\varepsilon_i c_{K_0}'X_{i(1)}'\left(\hat{F}_{(1)} - F_{(1)}^{*}\right)\right) \leq R_0(NT)^{-1}\left\|\hat{\Sigma}_{\hat{F}_{(1)}}^{-1}\right\|\left\|H_{(1)}\right\|\bar{A}_{NT}
$$

where $\bar{A}_{NT} = N^{-1}T^{-1}\left\|\sum_{i=1}^{N}F^{0\prime}\varepsilon_i c_{K_0}'X_{i(1)}'\left(\hat{F}_{(1)} - F_{(1)}^{*}\right)\right\|$. Noting that $\left\|H_{(1)}\right\| = O_P(1)$, and $\hat{\Sigma}_{\hat{F}_{(1)}}$ is asymptotically nonsingular by Lemma A.2(v) and Assumption A.1(ii), it suffices to show that $\bar{A}_{NT} = O_P(K_0^{-1/2}T^{1/2}\delta_{NT}^{-2})$. By (C.2), $\frac{1}{NT}\left\|\sum_{i=1}^{N}F^{0\prime}\varepsilon_i c_{K_0}'X_{i(1)}'(\hat{F}_{(1)} - F_{(1)}^{*})\right\| = \sum_{l=1}^{8}\frac{1}{NT}||\sum_{i=1}^{N}F^{0\prime}\varepsilon_i c_{K_0}'X_{i(1)}'a_{l(1)}|| \equiv \sum_{l=1}^{8}A_{lNT}$, say. By Assumptions A.1(iv), (v) and (vii),

$$
\begin{aligned}
A_{1NT} &= \frac{1}{(NT)^2}\left\|\sum_{i=1}^{N}F^{0\prime}\varepsilon_i c_{K_0}'X_{i(1)}'\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\hat{F}_{(1)}\right\| \leq \frac{R_0}{(NT)^2}\nu_{1NT}\|\boldsymbol{\varepsilon}\|_{\mathrm{sp}}^2\left\|\hat{F}_{(1)}\right\| \\
&= (NT)^{-2}O_P(NT)\,O_P(N+T)\,O_P\left(T^{1/2}\right) = O_P(N^{-1}T^{1/2} + T^{-1/2}),
\end{aligned}
$$

where $\nu_{1NT} = \left\| \sum_{i=1}^{N} F^{0\prime}\varepsilon_i c'_{K_0} X'_{i(1)} \right\|$ and we use the fact that

$$
\begin{aligned}
\nu_{1NT}^2 &= \sum_{i=1}^{N}\sum_{j=1}^{N} c'_{K_0} X'_{i(1)} X_{j(1)} c_{K_0} \varepsilon'_j F^0 F^{0\prime} \varepsilon_i \\
&\leq \left(\sum_{i=1}^{N}\sum_{j=1}^{N} c'_{K_0} X'_{i(1)} X_{j(1)} c_{K_0} c'_{K_0} X'_{j(1)} X_{i(1)} c_{K_0} \right)^{1/2} \left(\sum_{i=1}^{N}\sum_{j=1}^{N} \varepsilon'_j F^0 F^{0\prime} \varepsilon_i \varepsilon'_i F^0 F^{0\prime} \varepsilon_j \right)^{1/2} \\
&\leq \left(\sum_{i=1}^{N}\sum_{j=1}^{N} c'_{K_0} X'_{i(1)} X_{j(1)} X'_{j(1)} X_{i(1)} c_{K_0} \right)^{1/2} \sum_{i=1}^{N} \left\| F^{0\prime}\varepsilon_i \right\|^2 \\
&\leq \left[\mu_{\max}\left(\sum_{j=1}^{N} X_{j(1)} X'_{j(1)} \right) \mu_{\max}\left(\sum_{i=1}^{N} X'_{i(1)} X_{i(1)} \right) \right]^{1/2} \sum_{i=1}^{N} \left\| F^{0\prime}\varepsilon_i \right\|^2 = O_P(NT)\, O_P(NT).
\end{aligned}
$$

Let $H_B \equiv H_{(1)} V_{NT,11}^{-1}$, which corresponds to the H matrix in Bai (2009). By the triangle inequality,

$$
\begin{aligned}
A_{2NT} &= (NT)^{-2} \left\| \sum_{i=1}^{N} F^{0\prime}\varepsilon_i c'_{K_0} X'_{i(1)} F^0 \lambda^{0\prime} \boldsymbol{\varepsilon} \tilde{F}_{(1)} \right\| \\
&\leq (NT)^{-2} \left\| \sum_{i=1}^{N} F^{0\prime}\varepsilon_i c'_{K_0} X'_{i(1)} F^0 \lambda^{0\prime} \boldsymbol{\varepsilon} F^0 H_B \right\| + (NT)^{-2} \left\| \sum_{i=1}^{N} F^{0\prime}\varepsilon_i c'_{K_0} X'_{i(1)} F^0 \lambda^{0\prime} \boldsymbol{\varepsilon} \left(\tilde{F}_{(1)} - F^0 H_B \right) \right\| \\
&\equiv A_{2NT,1} + A_{2NT,2}, \text{ say.}
\end{aligned}
$$

By Assumptions A.1(ii) and (vii) and the fact $\|H_B\| = O_P(1)$,

$$
\begin{aligned}
A_{2NT,1} &\leq (NT)^{-2} \|\nu_{1NT}\| \left\|F^0\right\| \left\|\lambda^{0\prime}\boldsymbol{\varepsilon} F^0\right\| \|H_B\| \\
&= (NT)^{-2} O_P(NT)\, O_P\left(T^{1/2}\right) O_P\left(N^{1/2}T^{1/2}\right) O_P(1) = O_P(N^{-1/2}).
\end{aligned}
$$

Similarly, by Lemma A.2(ii) and Assumptions A.1(ii), (iii) and (v)

$$
\begin{aligned}
A_{2NT,2} &\leq (NT)^{-2} \|\nu_{1NT}\| \left\|F^0\lambda^{0\prime}\boldsymbol{\varepsilon}\right\| \left\| \tilde{F}_{(1)} - F^0 H_B \right\| \\
&= (NT)^{-2} O_P(NT)\, O_P\left(N^{1/2}T^{1/2}(N^{1/2}+T^{1/2})\right) O_P\left(T^{1/2}\delta_{NT}^{-1}\right) = O_P(T^{1/2}\delta_{NT}^{-2}).
\end{aligned}
$$

It follows that $A_{2NT} = O_P(T^{1/2}\delta_{NT}^{-2})$. By Assumption A.4(iii),

$$
\begin{aligned}
A_{3NT} &= (NT)^{-2} \left\| \sum_{i=1}^{N} F^{0\prime}\varepsilon_i c'_{K_0} X'_{i(1)} \boldsymbol{\varepsilon}' \lambda^0 F^{0\prime} \tilde{F}_{(1)} \right\| \leq (NT)^{-2} \left\| \sum_{i=1}^{N} F^{0\prime}\varepsilon_i c'_{K_0} X'_{i(1)} \boldsymbol{\varepsilon}' \lambda^0 \right\| \left\| F^{0\prime} \tilde{F}_{(1)} \right\| \\
&= (NT)^{-2} O_P\left(K_0^{1/2} NT \left(N^{1/2} + T^{1/2} \right) \right) O_P(T) = O_P\left(K_0^{1/2} T^{1/2} \delta_{NT}^{-2} \right).
\end{aligned}
$$

as

$$
\begin{aligned}
E\left| \sum_{i=1}^{N} c'_{R_0} F^{0\prime}\varepsilon_i c'_{K_0} X'_{i(1)} \boldsymbol{\varepsilon}' \lambda^0 \right|^2 &= E\left| \sum_{k=1}^{K_0} c_{kK_0} c'_{R_0} F^{0\prime} \boldsymbol{\varepsilon}' \mathbf{X}_k \boldsymbol{\varepsilon}' \lambda^0 \right|^2 \\
&\leq K_0 \sum_{k=1}^{K_0} c_{kK_0}^2 E\left(c'_{R_0} F^{0\prime} \boldsymbol{\varepsilon}' \mathbf{X}_k \boldsymbol{\varepsilon}' \lambda^0 \right)^2 = O\left(K_0 N^2 T^2 (N+T) \right).
\end{aligned}
$$

For the other terms, following the arguments as used in the analysis of Proposition B.1 and using the fact that $\nu_{1NT} = O_P(NT)$, we can readily apply Assumption A.1 to show that $A_{4NT} = O_P(KN^{-1}T^{-1/2})$, $A_{sNT} = O_P(K^{1/2}N^{-1})$ for $s = 5, 6$, and $A_{sNT} = O_P(K^{1/2}(N^{-1/2}T^{-1/2} + N^{-1}))$ for $s = 7, 8$. Consequently, $\frac{1}{NT}||F^{0\prime}\boldsymbol{\varepsilon}'\mathbf{X}_k(\hat{F}_{(1)} - F^*_{(1)})|| = O_P(K_0^{1/2}T^{1/2}\delta_{NT}^{-2})$ and (ii) follows.

(iii) Note that

$$
\begin{aligned}
& \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} X'_{i(1)}F^*_{(1)}\left(\hat{F}'_{(1)}\hat{F}_{(1)}\right)^{-1}\left(\hat{F}_{(1)} - F^*_{(1)}\right)'\varepsilon_i \\
= & \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} X'_{i(1)}F^*_{(1)}\left(F^{*\prime}_{(1)}F^*_{(1)}\right)^{-1}\left(\hat{F}_{(1)} - F^*_{(1)}\right)'\varepsilon_i \\
& + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} X'_{i(1)}F^*_{(1)}\left[\left(\hat{F}'_{(1)}\hat{F}_{(1)}\right)^{-1} - \left(F^{*\prime}_{(1)}F^*_{(1)}\right)^{-1}\right]\left(\hat{F}_{(1)} - F^*_{(1)}\right)'\varepsilon_i \\
\equiv & J_1 + J_2, \text{ say.}
\end{aligned}
$$

By Lemmas A.2(i) and (iii), Assumptions A.1(iv)-(v), for any arbitrary $K_0 \times 1$ vector c_{K_0} with $\|c_{K_0}\| = 1$,

$$
\begin{aligned}
\left|c'_{K_0}J_2\right| = & \frac{1}{\sqrt{NT}}\left|\operatorname{tr}\left(F^*_{(1)}\left[\left(\hat{F}'_{(1)}\hat{F}_{(1)}\right)^{-1} - \left(F^{*\prime}_{(1)}F^*_{(1)}\right)^{-1}\right]\left(\hat{F}_{(1)} - F^0_{(1)}\right)'\sum_{k=1}^{K_0} c_{kK_0}\boldsymbol{\varepsilon}'\mathbf{X}_k\right)\right| \\
\leq & R_0(NT)^{-1/2}\left\{\frac{1}{T}\left\|F^*_{(1)}\right\|\left\|\hat{F}_{(1)} - F^*_{(1)}\right\|\right\}\left\|\left(T^{-1}\hat{F}'_{(1)}\hat{F}_{(1)}\right)^{-1} - \left(T^{-1}F^{*\prime}_{(1)}F^*_{(1)}\right)^{-1}\right\|\|\boldsymbol{\varepsilon}\|_{\text{sp}} \\
& \times\left\{\sum_{k=1}^{K_0}\|\mathbf{X}_k\|^2_{\text{sp}}\right\}^{1/2} \\
= & (NT)^{-1/2}O_P(\delta_{NT}^{-1})O_P(\delta_{NT}^{-2} + (NT/K)^{-1/2})O_P(N^{1/2} + T^{1/2})O_P((K_0NT)^{1/2}) \\
= & O_P\left(K^{1/2}K_0^{1/2}\delta_{NT}^{-2}\right).
\end{aligned}
$$

Hence $\|J_2\| = O_P(K^{1/2}K_0^{1/2}\delta_{NT}^{-2})$. For J_1, by (C.2) we have

$$
\begin{aligned}
& \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} X'_{i(1)}F^*_{(1)}\left(F^{*\prime}_{(1)}F^*_{(1)}\right)^{-1}\left(\hat{F}_{(1)} - F^*_{(1)}\right)'\varepsilon_i \\
= & \sum_{l=1}^{8}\frac{1}{N^{1/2}T^{1/2}}\sum_{i=1}^{N} X'_{i(1)}F^*_{(1)}\left(F^{*\prime}_{(1)}F^*_{(1)}\right)^{-1}a'_{l(1)}\varepsilon_i \equiv \sum_{l=1}^{8} A_{NT}(l), \text{ say.}
\end{aligned}
$$

One can readily show that $A_{NT}(l) = O_P(K_0^{1/2}\delta_{NT}^{-1})$ for $l = 4, 5, ..., 8$. For $A_{NT}(1)$, we apply Assumptions A.1(iv)-(v) to obtain

$$
\begin{aligned}
\left|c'_{K_0}A_{NT}(1)\right| = & \frac{1}{N^{3/2}T^{3/2}}\left|\operatorname{tr}\left(F^*_{(1)}\left(F^{*\prime}_{(1)}F^*_{(1)}\right)^{-1}\tilde{F}'_{(1)}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\sum_{k=1}^{K_0} c_{kK_0}\boldsymbol{\varepsilon}'\mathbf{X}_k\right)\right| \\
\leq & \frac{R_0}{N^{3/2}T^{3/2}}\left\|F^*_{(1)}\left(F^{*\prime}_{(1)}F^*_{(1)}\right)^{-1}\tilde{F}'_{(1)}\right\|_{\text{sp}}\|\boldsymbol{\varepsilon}\|^3_{\text{sp}}\left\{\sum_{k=1}^{K_0}\|\mathbf{X}_k\|^2_{\text{sp}}\right\}^{1/2} \\
= & (NT)^{-3/2}O_P(1)O_P(N^{3/2} + T^{3/2})O_P((K_0NT)^{1/2}) = O_P(K_0^{1/2}\delta_{NT}^{-1}),
\end{aligned}
$$

where we also use the fact that $|\mathrm{tr}(A)| \leq \mathrm{rank}(A) \|A\|_{\mathrm{sp}}$, and the submultiplicative property of the spectral norm, and the triangle inequality. Hence $\|A_{NT}(1)\| = O_P(K_0^{1/2}\delta_{NT}^{-1})$. Next, we decompose $A_{NT}(2)$ as follows:

$$\begin{aligned} A_{NT}(2) &= \frac{1}{N^{3/2}T^{3/2}} \sum_{i=1}^{N} X_{i(1)}' F_{(1)}^* \left(F_{(1)}^{*\prime} F_{(1)}^*\right)^{-1} H_B' F^{0\prime} \boldsymbol{\varepsilon}' \lambda^{0\prime} F^{0\prime} \varepsilon_i \\ &\quad + \frac{1}{N^{3/2}T^{3/2}} \sum_{i=1}^{N} X_{i(1)}' F_{(1)}^* \left(F_{(1)}^{*\prime} F_{(1)}^*\right)^{-1} \left(\tilde{F}_{(1)} - F^0 H_B\right)' \boldsymbol{\varepsilon}' \lambda^{0\prime} F^{0\prime} \varepsilon_i \\ &\equiv A_{NT}(2,1) + A_{NT}(2,2), \text{ say.} \end{aligned}$$

By Assumptions A.1(ii), (iv), (v) and (vii),

$$\begin{aligned} \left|c_{K_0}' A_{NT}(2,1)\right| &= \left| \frac{1}{N^{3/2}T^{3/2}} \sum_{i=1}^{N} c_{K_0}' X_{i(1)}' F_{(1)}^* \left(F_{(1)}^{*\prime} F_{(1)}^*\right)^{-1} H_B' F^{0\prime} \boldsymbol{\varepsilon}' \lambda^{0\prime} F^{0\prime} \varepsilon_i \right| \\ &= \frac{1}{N^{3/2}T^{3/2}} \left| \mathrm{tr}\left(F_{(1)}^* \left(F_{(1)}^{*\prime} F_{(1)}^*\right)^{-1} H_B' F^{0\prime} \boldsymbol{\varepsilon}' \lambda^{0\prime} F^{0\prime} \sum_{k=1}^{K_0} c_{kK_0} \boldsymbol{\varepsilon}' \mathbf{X}_k \right) \right| \\ &\leq \frac{R_0}{N^{3/2}T^{3/2}} \left\| F_{(1)}^* \left(F_{(1)}^{*\prime} F_{(1)}^*\right)^{-1} H_B' \right\| \left\| F^{0\prime} \boldsymbol{\varepsilon}' \lambda^{0\prime} \right\| \left\| F^0 \right\| \|\boldsymbol{\varepsilon}\|_{\mathrm{sp}} \left\{ \sum_{k=1}^{K_0} \|\mathbf{X}_k\|^2 \right\}^{1/2} \\ &= (NT)^{-3/2} O_P\left(T^{-1/2}\right) O_P\left(N^{1/2}T^{1/2}\right) O_P\left(T^{1/2}\right) O_P\left(N^{1/2} + T^{1/2}\right) O_P\left((K_0 NT)^{1/2}\right) \\ &= O_P(K_0^{1/2}\delta_{NT}^{-1}). \end{aligned}$$

Similarly,

$$\begin{aligned} \left|c_{K_0}' A_{NT}(2,2)\right| &= \left| \frac{1}{N^{3/2}T^{3/2}} \sum_{i=1}^{N} c_{R_0}' X_{i(1)}' F_{(1)}^* \left(F_{(1)}^{*\prime} F_{(1)}^*\right)^{-1} \left(\tilde{F}_{(1)} - F^0 H_B\right)' \boldsymbol{\varepsilon}' \lambda^{0\prime} F^{0\prime} \varepsilon_i \right| \\ &= \frac{1}{N^{3/2}T^{3/2}} \left| \mathrm{tr}\left(F_{(1)}^* \left(F_{(1)}^{*\prime} F_{(1)}^*\right)^{-1} \left(\tilde{F}_{(1)} - F^0 H_B\right)' \boldsymbol{\varepsilon}' \lambda^{0\prime} F^{0\prime} \sum_{k=1}^{K_0} c_{kK_0} \boldsymbol{\varepsilon}' \mathbf{X}_k \right) \right| \\ &\leq \frac{R_0}{N^{3/2}T^{3/2}} \left\| F_{(1)}^* \left(F_{(1)}^{*\prime} F_{(1)}^*\right)^{-1} \right\| \left\| \tilde{F}_{(1)} - F^0 H_B \right\| \left\| \boldsymbol{\varepsilon}' \lambda^{0\prime} \right\| \left\| F^0 \right\| \|\boldsymbol{\varepsilon}\|_{\mathrm{sp}} \left\{ \sum_{k=1}^{K_0} \|\mathbf{X}_k\|^2 \right\}^{1/2} \\ &= (NT)^{-3/2} O_P\left(T^{-1/2}\right) O_P\left(T^{1/2}\delta_{NT}^{-1}\right) O_P\left(N^{1/2}T^{1/2}\right) O_P\left(T^{1/2}\right) O_P\left(N^{1/2} + T^{1/2}\right) \\ &\quad \times O_P\left((K_0 NT)^{1/2}\right) \\ &= O_P\left(K_0^{1/2} T^{1/2} \delta_{NT}^{-2}\right). \end{aligned}$$

It follows that $A_{NT}(2) = O_P(K_0^{1/2}T^{1/2}\delta_{NT}^{-2})$. For $A_{NT}(3)$, we have

$$A_{NT}(3) = \frac{1}{N^{3/2}T^{3/2}} \sum_{i=1}^{N} X_{i(1)}' F_{(1)}^* \left(F_{(1)}^{*\prime} F_{(1)}^*\right)^{-1} \tilde{F}_{(1)}' F^0 \lambda^{0\prime} \boldsymbol{\varepsilon} \varepsilon_i = \mathbb{B}_{3NT}.$$

This completes the proof of (iii).

(iv) By (C.2), $\frac{1}{T}(\hat{F}_{(1)} - F_{(1)}^0)' \varepsilon_i = \sum_{l=1}^{8} \frac{1}{T} a_{l(1)}' \varepsilon_i = \sum_{l=1}^{8} A_{NT}(l)$, say. The proof is analogous to that of Lemma A.2(iii) and we only show that $A_{NT}(l) = O_P\left(\delta_{NT}^{-2}\right)$ for $l = 1, 2, 3$ as the analysis of the other terms is simpler. By the triangle inequality, the relationship between the spectral and Frobenius norms and their submultiplicative property, and Assumptions A.1(ii), (v), and (vii), A.2(i), and A.4(iv),

$$\begin{aligned}\|A_{NT}(1)\| &= N^{-1}T^{-2}\left\|\tilde{F}_{(1)}'\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\varepsilon_i\right\| \le R_0 N^{-1}T^{-2}\left\|\tilde{F}_{(1)}\right\|\|\boldsymbol{\varepsilon}\|_{\text{sp}}^2\|\varepsilon_i\| \\ &= N^{-1}T^{-2}O_P\left(T^{1/2}\right)O_P(N+T)\,O_P\left(T^{1/2}\right) = O_P\left(\delta_{NT}^{-2}\right), \\ \|A_{NT}(2)\| &= N^{-1}T^{-2}\left\|\tilde{F}_{(1)}'\boldsymbol{\varepsilon}'\lambda^0F^{0\prime}\varepsilon_i\right\| \le N^{-1}T^{-2}\left\|\tilde{F}_{(1)}\right\|\|\boldsymbol{\varepsilon}'\lambda^0\|\|F^{0\prime}\varepsilon_i\| \\ &= N^{-1}T^{-2}O_P\left(T^{1/2}\right)O_P\left(N^{1/2}T^{1/2}\right)O_P\left(T^{1/2}\right) = O_P\left(N^{-1/2}T^{-1/2}\right),\end{aligned}$$

and

$$\begin{aligned}\|A_{NT}(3)\| &= N^{-1}T^{-2}\left\|\tilde{F}_{(1)}'F^0\lambda^{0\prime}\boldsymbol{\varepsilon}\varepsilon_i\right\| \le N^{-1}T^{-2}\left\|\tilde{F}_{(1)}'F^0\right\|\|\lambda^{0\prime}\boldsymbol{\varepsilon}\varepsilon_i\| \\ &= N^{-1}T^{-2}O_P(T)\,O_P\left(N^{1/2}T^{1/2}+T\right) = O_P\left(N^{-1}+N^{-1/2}T^{-1/2}\right).\end{aligned}$$

This completes the proof of (iv). ■

D Some primitive assumptions and technical lemmas

In this appendix we present two assumptions that replace some high level conditions in Assumptions A1, A2, A4 and A5 in the text. They are also used to prove the technical lemmas in this appendix and Corollary 3.4 in the next appendix.

Recall that $\mathcal{D} \equiv \sigma\left(F^0, \lambda^0\right)$ and $E_{\mathcal{D}}(A) = E(A|\mathcal{D})$. Let $\|A\|_{q,\mathcal{D}} \equiv [E_{\mathcal{D}}(\|A\|_F^q)]^{1/q}$.

Assumption B.1 (i) $\max_{1\le t\le T} E||F_t^0||^{8+4\sigma} \le C$ for some $\sigma > 0$ and $C < \infty$ and $T^{-1}F^{0\prime}F^0 \xrightarrow{P} \Sigma_{F^0} > 0$ as $T \to \infty$.

(ii) $\max_{1\le i\le N} E||\lambda_i^0||^{8+4\sigma} \le C$ and $N^{-1}\lambda^{0\prime}\lambda^0 \xrightarrow{P} \Sigma_{\lambda^0} > 0$ as $N \to \infty$.

(iii) $\max_{1\le i\le N, 1\le t\le T} E\left|\varepsilon_{it}\right|^{8+4\sigma}$ and $\max_{1\le k\le K}\max_{1\le i\le N, 1\le t\le T} E\left\|X_{it,k}\right\|^{8+4\sigma} \le C$.

(iv) $\max_{1\le i\le N} T^{-1}\sum_{t=1}^T E_{\mathcal{D}}\left(\varepsilon_{it}^2\right) = O_P(1)$ and $\max_{1\le t\le T} N^{-1}\sum_{i=1}^N E_{\mathcal{D}}\left(\varepsilon_{it}^2\right) = O_P(1)$, $\max_{1\le k\le K_0}\max_{1\le i\le N} T^{-1}\sum_{t=1}^T E_{\mathcal{D}}(X_{it,k}^2) = O_P(1)$ and $\max_{1\le k\le K_0}\max_{1\le t\le T} N^{-1}\sum_{i=1}^N E_{\mathcal{D}}(X_{it,k}^2) = O_P(1)$.

(v) $\max_{1\le t\le T} N^{-1}\sum_{i=1}^N \|\varepsilon_{it}\|_{8+4\sigma,\mathcal{D}}^4 = O_P(1)$ and $\max_{1\le k\le K_0}\max_{1\le t\le T} N^{-1}\sum_{i=1}^N \|X_{it,k}\|_{8+4\sigma,\mathcal{D}}^4 = O_P(1)$.

Note that Assumptions B.1(i)-(iii) strengthen the moment conditions in Assumptions A.1(ii)-(iv) and A.2(i) and require finite eighth plus moments for F_t^0, λ_i^0, X_{it}, and ε_{it} to derive the asymptotic distribution of our adaptive group Lasso estimator and to estimate the asymptotic bias and variance terms. Admittedly, our moment conditions are generally different and may sometimes be stronger than those assumed in the literature (e.g., Bai, 2009). For example, Bai (2009) only requires finite fourth moments for F_t^0, λ_i^0 and X_{it} and finite eighth moments for ε_{it}; but he assumes independence between ε_{it} and $(X_{js}, F_s^0, \lambda_j^0)$ for all i,j,t,s, and thus rules out dynamics in the model. Moon and Weidner

(2013) assume eighth moments for ε_{it}; but they also assume that both the factors and factor loadings are uniformly bounded. In addition, they assume that the error terms ε_{it} are independent across both i and t, which may rule out conditional heteroskedasticity in dynamic panels. Assumptions B.1(iv)-(v) are needed to show some uniform results below.

To state the next assumption, we first provide the definition of conditional strong mixing processes.

Definition D.1 *Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. Let $P_{\mathcal{B}}(\cdot) \equiv P(\cdot|\mathcal{B})$. Let $\{\xi_t, t \geq 1\}$ be a sequence of random variables defined on $(\Omega, \mathcal{A}, P)$. The sequence $\{\xi_t, t \geq 1\}$ is said to be conditionally strong mixing given $\mathcal{B}$ (or $\mathcal{B}$-strong-mixing) if there exists a nonnegative $\mathcal{B}$-measurable random variable $\alpha^{\mathcal{B}}(t)$ converging to 0 a.s. as $t \to \infty$ such that*

$$|P_{\mathcal{B}}(A \cap B) - P_{\mathcal{B}}(A) P_{\mathcal{B}}(B)| \leq \alpha^{\mathcal{B}}(t) \ \ a.s. \tag{D.1}$$

for all $A \in \sigma(\xi_1, ..., \xi_k)$, $B \in \sigma(\xi_{k+t}, \xi_{k+t+1}, ...)$ and $k \geq 1$, $t \geq 1$.

The above definition is due to Prakasa Rao (2009); see also Roussas (2008). When one takes $\alpha^{\mathcal{B}}(t)$ as the supremum of the left hand side object in (D.1) over the set $\{A \in \sigma(\xi_1, ..., \xi_k), B \in \sigma(\xi_{k+t}, \xi_{k+t+1}, ...), k \geq 1\}$, we refer to it as the $\mathcal{B}$-strong-mixing coefficient.

Assumption B.2 (i) For each $i = 1, ..., N$, $\{(X_{it}, \varepsilon_{it}) : t = 1, 2, ...\}$ is conditionally strong mixing given $\mathcal{D}$ with mixing coefficients $\{\alpha^{\mathcal{D}}_{NT,i}(\cdot)\}$. $\alpha_{\mathcal{D}}(\cdot) \equiv \alpha^{\mathcal{D}}_{NT}(\cdot) \equiv \max_{1 \leq i \leq N} \alpha^{\mathcal{D}}_{NT,i}(\cdot)$ satisfies $\alpha_{\mathcal{D}}(s) = O_{a.s.}(s^{-\rho})$ where $\rho = (2+\sigma)/(1+\sigma) + \epsilon$ for some arbitrarily small $\epsilon > 0$ and σ is as defined in Assumption B.1(i). In addition, there exist integers τ_0, $\tau_* \in (1, T)$ such that $NT\alpha_{\mathcal{D}}(\tau_0) = o_{a.s.}(1)$, $T(T + N^{1/2})\alpha_{\mathcal{D}}(\tau_*)^{(1+\sigma)/(2+\sigma)} = o_{a.s.}(1)$, and $N^{1/2}T^{-1}\tau_*^2 = o(1)$.

(ii) (ε_i, X_i), $i = 1, ..., N$, are mutually independent of each other conditional on $\mathcal{D}$.

(iii) For each $i = 1, ..., N$, $E(\varepsilon_{it}|\mathcal{F}_{NT,t-1}) = 0$ a.s., where $\mathcal{F}_{NT,t} \equiv \sigma(\mathcal{D}, \{X_{i,t+1}, X_{it}, \varepsilon_{it}, X_{i,t-1}, \varepsilon_{i,t-1}, ...\}_{i=1}^{N})$.

(iv) As $(N, T) \to \infty$, $K_0^{1/2} M N^{1/2} T^{-1/2} (\delta_{NT}^{-1} + \alpha_{NT}) \to 0$ and $K_0 NT \alpha_{\mathcal{D}}(M+1)^{(3+2\sigma)/(2+\sigma)} = o_{a.s.}(1)$, where α_{NT} is defined in Lemma D.2 below.

B.2(i) requires that each individual time series $\{(X_{it}, \varepsilon_{it}) : t = 1, 2, ...\}$ be $\mathcal{D}$-strong-mixing. To appreciate the importance of conditioning, we take the simple panel AR(1) model considered by Su and Chen (2013) as an example:

$$Y_{it} = \rho_0 Y_{i,t-1} + \lambda_i^{0\prime} F_t^0 + \varepsilon_{it}, \ i = 1, ..., N, \ t = 1, \ldots, T. \tag{D.2}$$

Even if $\{(\varepsilon_{it}, F_t^0), t \geq 1\}$ is a strong mixing process, $\{Y_{it}, t \geq 1\}$ is generally not unless λ_i^0 is nonstochastic. For this reason, Hahn and Kuersteiner (2011) assume that the individual fixed effects are nonrandom and uniformly bounded in their study of nonlinear dynamic panel data models. In the case of random fixed effects, they suggest adopting the concept of conditional strong mixing where the mixing coefficient is defined by conditioning on the fixed effects. Our spirit is similar to theirs as we define the conditional strong mixing processes by conditioning on both factors and factor loadings. The dependence of the mixing rate on σ defined in B.1 reflects the trade-off between the degree of dependence and the moment bounds of the process $\{(X_{it}, \varepsilon_{it}), t \geq 1\}$. As Su and Chen (2013) remark, Assumption B.2(ii) does not rule out cross sectional dependence among $(X_{it}, \varepsilon_{it})$. When $X_{it} = Y_{i,t-1}$ and ε_{it} exhibits conditional heteroskedasticity (e.g., $\varepsilon_{it} = \sigma_0 (Y_{i,t-1}) \epsilon_{it}$ where $\epsilon_{it} \sim$IID$(0,1)$ and $\sigma_0 (\cdot)$ is an unknown smooth function) as in (D.2), $(X_{it}, \varepsilon_{it})$ are not independent across i because of the presence of common factors irrespective of whether one allows λ_i^0 to be independent across i or not. Nevertheless, conditional on $\mathcal{D}$, it is possible that $(X_{it}, \varepsilon_{it})$ is independent across i such that A.2(ii) is still satisfied. Note that here the cross sectional dependence is similar to the type of cross sectional dependence generated by common shocks studied by Andrews (2005), but the latter author assumes IID observations conditional on the σ-field generated by the common shocks in a cross-section framework. B.2(iii) requires that the error term ε_{it} be a martingale difference sequence (m.d.s.) with respect to the filter $\mathcal{F}_{NT,t}$ which allows for lagged dependent variables in X_{it}, and conditional heteroskedasticity, skewness, or kurtosis of an unknown form in ε_{it}. In contrast, both Bai (2009) and Pesaran (2006) assume that ε_{it} is independent of X_{js}, λ_j, and F_s for all i, t, j and s; Moon and Weidner (2013) allow dynamics but assume that ε_{it}'s are independent across both i and t. The allowance of lagged dependent variables broadens the potential applicability of our shrinkage estimation method. B.2(iv) requires that M should not grow too fast.

To proceed, we remark that with Assumptions B.1-B.2, the high level conditions in Assumptions A.1(vi)-(vii), A.2(ii)-(iii), A.4(iii)-(iv) and A.5(i)-(ii) can be easily verified. Assumption A.1(vi) follows from Assumptions B.1(iii) and B.2(ii)-(iii) and Chebyshev inequality. Assumptions A.1(vii) holds because $E \left\| F^{0\prime} \boldsymbol{\varepsilon}' \lambda^0 \right\|^2 = \sum_{1 \leq i,j \leq N} \sum_{1 \leq t,s \leq T} E[E_{\mathcal{D}} (\varepsilon_{it} \varepsilon_{js}) F_s^{0\prime} F_t^0 \lambda_i^{0\prime} \lambda_j^0] = \sum_{i=1}^N \sum_{t=1}^T E \left[E_{\mathcal{D}} \left(\varepsilon_{it}^2 \right) ||F_t^0||^2 ||\lambda_i^0||^2 \right]$ $= O(NT)$ under Assumptions B.1(i)-(iii) and B.2(ii)-(iii). Assumption A.2(ii) is trivially satisfied under Assumption B.1(iii) and B.2(ii)-(iii) and A.2(iii) can be verified under Assumptions B.1(iii) and B.2(ii) by the law of iterated expectations. Assumptions A.4(iii)-(iv) follow because we can show that

$$\begin{aligned} E_{\mathcal{D}} \left\| F^{0\prime} \boldsymbol{\varepsilon}' \mathbf{X}_k \boldsymbol{\varepsilon}' \lambda^0 \right\|^2 &= \sum_{1 \leq i,j,l,m \leq N} \sum_{1 \leq t,s,r,q \leq T} E_{\mathcal{D}} \left[\varepsilon_{it} X_{is,k} \varepsilon_{js} \varepsilon_{lr} X_{mr,k} \varepsilon_{mq} \right] F_q^{0\prime} F_t^0 \lambda_j^{0\prime} \lambda_l^0 \\ &= O_P \left(N^2 T^2 (N+T) \right) \text{ and} \\ E_{\mathcal{D}} \left\| \lambda^{0\prime} \boldsymbol{\varepsilon} \varepsilon_i \right\|^2 &= \sum_{j=1}^N \sum_{k=1}^N \sum_{t=1}^T \sum_{s=1}^T E_{\mathcal{D}} \left[\varepsilon_{jt} \varepsilon_{it} \varepsilon_{is} \varepsilon_{ks} \right] \lambda_k^{0\prime} \lambda_j^0 = O_P \left(T(N+T) \right) \end{aligned}$$

under Assumptions B.2(i)-(iii) by the use of Davydov inequality for conditional strong mixing processes; see, e.g., the proof of Lemma D.3(vi) below. Finally, Assumptions A.5(i)-(ii) follow under B.1-B.2 by straightforward verification of the moment conditions for the martingale central limit theorem (e.g., Pollard (1984, p.171)).

To prove Corollary 3.4, we need several lemmas.

Lemma D.2 *Suppose that the conditions in Corollary 3.4 hold. Then*
(i) $\max_{1\leq i\leq N}\left|T^{-1}\sum_{t=1}^{T}\left[\varepsilon_{it}^{2}-E_{\mathcal{D}}\left(\varepsilon_{it}^{2}\right)\right]\right|=O_{P}\left(\alpha_{NT}\right);$
(ii)$\max_{1\leq i\leq N}T^{-1}\left\|\varepsilon_{i}\right\|^{2}=O_{P}\left(1\right);$
(iii) $\max_{1\leq i\leq N}\left\|T^{-1}F^{0\prime}\varepsilon_{i}\right\|=O_{P}\left(\alpha_{NT}\right);$
(iv) $\max_{1\leq i\leq N}\left\|T^{-1}X_{i(1)}^{\prime}\varepsilon_{i}\right\|=O_{P}(K_{0}^{1/2}\alpha_{NT});$
(v) $\max_{1\leq i\leq N}T^{-1}\left\|X_{i(1)}\right\|^{2}=O_{P}\left(K_{0}\right);$
(vi) $\max_{1\leq i\leq N}\left\|\lambda_{i(1)}^{*}\right\|=o_{P}\left(N^{1/(8+4\sigma)}\right);$
(vii) $\max_{1\leq i\leq N}T^{-1}\left\|\left[X_{i(1)}^{\prime}-E_{\mathcal{D}}(X_{i(1)}^{\prime})\right]F^{0}\right\|=O_{P}(K_{0}^{1/2}\alpha_{NT});$
where $\alpha_{NT}=\max\{\left(NT\right)^{1/(4+2\sigma)}\log\left(NT\right)/T,\ (\log\left(NT\right)/T)^{1/2}\}.$

Proof. (i) The proof is analogous to that of Lemma A.7(iii) in Su and Chen (2013) by using Bernstein inequality for conditional strong mixing processes (see, e.g., Lemma A.4 in Su and Chen (2013)).

(ii)$\max_{1\leq i\leq N}T^{-1}\left\|\varepsilon_{i}\right\|^{2}\leq\max_{1\leq i\leq N}\left|T^{-1}\sum_{t=1}^{T}\left[\varepsilon_{it}^{2}-E_{\mathcal{D}}\left(\varepsilon_{it}^{2}\right)\right]\right|+\max_{1\leq i\leq N}\left|T^{-1}\sum_{t=1}^{T}E_{\mathcal{D}}\left(\varepsilon_{it}^{2}\right)\right|=O_{P}\left(\alpha_{NT}\right)+O_{P}\left(1\right)=O_{P}\left(1\right)$ by (i) and Assumption B.1(iv).

(iii) The proof is analogous to that of (i).

(iv) The proof is analogous to that of (i).

(v) Following the proof of (i), we can show that $\max_{1\leq i\leq N}\left|T^{-1}\sum_{t=1}^{T}X_{it(1)}^{\prime}X_{it(1)}-E_{\mathcal{D}}(X_{it(1)}^{\prime}X_{it(1)})\right|=O_{P}\left(K_{0}\alpha_{NT}\right).$ Then the result follows from this, Assumption B.1(iv) and the triangle inequality.[19]

(vi) By Boole and Markov inequalities, for any $\epsilon>0$ we have

$$\begin{aligned}P\left(\max_{1\leq i\leq N}\left\|\lambda_{i}^{0}\right\|\geq N^{1/(8+4\sigma)}\epsilon\right) &\leq N\max_{1\leq i\leq N}P\left(\left\|\lambda_{i}^{0}\right\|\geq N^{1/(8+4\sigma)}\epsilon\right)\\ &\leq \epsilon^{-(8+4\sigma)}\max_{1\leq i\leq N}E\left\{\left\|\lambda_{i}^{0}\right\|^{8+4\delta}1\left\{\left\|\lambda_{i}^{0}\right\|\geq N^{1/(8+4\sigma)}\epsilon\right\}\right\}\\ &\rightarrow 0,\end{aligned}$$

where the last line follows from Assumption B.1(ii) and the dominated convergence theorem. It follows that $\max_{1\leq i\leq N}\left\|\lambda_{i}^{0}\right\|=o_{P}(N^{1/(8+4\sigma)})$. The conclusion follows as one can write $\lambda_{i(1)}^{*}=H_{(1)}^{-1}\lambda_{i}^{0}$ and $H_{(1)}$ is asymptotically nonsingular.

(vii) The proof is analogous to that of (i). ■

Let $\Psi_{NT}\equiv$diag$(\psi_{1T},...,\psi_{NT})$ and $\psi_{iT}\equiv T^{-1}\sum_{t=1}^{T}E_{\mathcal{D}}\left[\varepsilon_{it}^{2}\right].$ Recall $\hat{\Psi}_{NT}=$diag$(\hat{\psi}_{1T},...,\hat{\psi}_{NT})$ and $\hat{\psi}_{iT}=T^{-1}\hat{\varepsilon}_{i}^{\prime}\hat{\varepsilon}_{i}=T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^{2}$, where $\hat{\varepsilon}_{i}\equiv(\hat{\varepsilon}_{i1},...,\hat{\varepsilon}_{iT})^{\prime}$.

Lemma D.3 *Suppose that the conditions in Corollary 3.4 hold. Then*
(i) $\lambda^{0\prime}\Psi_{NT}\lambda^{0}=O_{P}\left(N\right);$
(ii) $\lambda^{0\prime}\left(T^{-1}\boldsymbol{\varepsilon\varepsilon}^{\prime}-\Psi_{NT}\right)\lambda^{0}=O_{P}\left(N^{1/2}+NT^{-1/2}\right);$
(iii) $\left\|\hat{\Psi}_{NT}-\Psi_{NT}\right\|_{sp}=\max_{1\leq i\leq N}\left|\hat{\psi}_{iT}-\psi_{iT}\right|=o_{P}\left(\delta_{NT}^{-1}N^{1/(8+4\sigma)}\right);$

[19] Alternatively, under Assumption B.1(iii), the fourth moment of $T^{-1}\left\|X_{i(1)}\right\|^{2}=T^{-1}\sum_{t=1}^{T}X_{it(1)}^{\prime}X_{it(1)}$ is finite. Then following the proof of (vi) below, we have $\max_{1\leq i\leq N}T^{-1}\left\|X_{i(1)}\right\|^{2}=o_{P}(K_{0}^{1/2}N^{1/4})$, a rough bound that also suffices for our purpose, but stringent conditions are required on the relative rates at which K_{0}, N and T pass to infinity.

(iv) $(NT)^{-1}\left\|\hat{F}_{(1)}\hat{\lambda}'\hat{\Psi}_{NT}\hat{\lambda}\hat{F}_{(1)}'-F^0\lambda^{0\prime}\Psi_{NT}\lambda^0F^{0\prime}\right\|=o_P\left(\delta_{NT}^{-1}N^{1/(8+4\sigma)}\right)$;

(v) $(NT)^{-1}\left(\hat{\lambda}'\hat{\Psi}_{NT}\mathbf{X}_k\hat{F}_{(1)}-H^{+}\lambda^{0\prime}\Psi_{NT}\mathbf{X}_kF_{(1)}^{*}\right)=O_P\left(\delta_{NT}^{-1}\right)$ *for* $k=1,...,K_0$;

(vi) $N^{-3}E_{\mathcal{D}}\left\|\lambda^{0\prime}\left(T^{-1}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'-\Psi_{NT}\right)\mathbf{X}_k\right\|^2=O_P\left(N^{-1}+T^{-1}\right)$ *for* $k=1,...,K_0$.

Proof. (i) By Assumption B.1(ii) and (iv) $\lambda^{0\prime}\Psi_{NT}\lambda^0\leq\max_{1\leq i\leq N}\psi_{iT}\left\|\lambda^0\right\|^2=O_P(1)\,O_P(N)=O_P(N)$.

(ii) Under Assumption B.1(iv), $T^{-1}E_{\mathcal{D}}\left[\lambda^{0\prime}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\lambda^0\right]=\sum_{i=1}^N\lambda_i^0\lambda_i^{0\prime}T^{-1}\sum_{t=1}^TE_{\mathcal{D}}\left[\varepsilon_{it}^2\right]=\lambda^{0\prime}\Psi_{NT}\lambda^0$. Let $\Xi_{NT}=c_{R_0}'\left(T^{-1}\lambda^{0\prime}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\lambda^0-\lambda^{0\prime}\Psi_{NT}\lambda^0\right)\bar{c}_{R_0}$ where $\bar{c}_{R_0}$ is similarly defined as c_{R_0} with $\|\bar{c}_{R_0}\|=1$. Then $E_{\mathcal{D}}(\Xi_{NT})=0$. Noting that by Assumptions B.2(ii)-(iii)

$$T^{-2}\sum_{t=1}^T\sum_{s=1}^TE_{\mathcal{D}}\left[\varepsilon_{it}\varepsilon_{jt}\varepsilon_{ks}\varepsilon_{ls}\right]=\begin{cases}\psi_{iT}\psi_{kT} & \text{if } i=j\neq k=l\\ T^{-2}\sum_{t=1}^TE_{\mathcal{D}}\left[\varepsilon_{it}^2\varepsilon_{jt}^2\right] & \text{if } i=k\neq j=l \text{ or } i=l\neq j=k\\ T^{-2}\sum_{t=1}^T\sum_{s=1}^TE_{\mathcal{D}}\left[\varepsilon_{it}^2\varepsilon_{is}^2\right] & \text{if } i=j=k=l\\ 0 & \text{otherwise}\end{cases}\tag{D.3}$$

we have

$$\begin{aligned}\mathrm{Var}_{\mathcal{D}}(\Xi_{NT}) &= E_{\mathcal{D}}\left[\left(c_{R_0}'T^{-1}\lambda^{0\prime}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\lambda^0\bar{c}_{R_0}\right)^2\right]-\left[c_{R_0}'\lambda^{0\prime}\Psi_{NT}\lambda^0\bar{c}_{R_0}\right]^2\\ &= T^{-2}E_{\mathcal{D}}\mathrm{tr}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\lambda^0\bar{c}_{R_0}c_{R_0}'\lambda^{0\prime}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\lambda^0\bar{c}_{R_0}c_{R_0}'\lambda^{0\prime}\right]-\left[c_{R_0}'\lambda^{0\prime}\Psi_{NT}\lambda^0\bar{c}_{R_0}\right]^2\\ &= T^{-2}\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^N\sum_{l=1}^N\sum_{t=1}^T\sum_{s=1}^Tc_{R_0}'\lambda_i^0\lambda_j^{0\prime}\bar{c}_{R_0}c_{R_0}'\lambda_k^0\lambda_l^{0\prime}\bar{c}_{R_0}E_{\mathcal{D}}\left[\varepsilon_{it}\varepsilon_{jt}\varepsilon_{ks}\varepsilon_{ls}\right]-\left[\sum_{i=1}^Nc_{R_0}'\lambda_i^0\lambda_j^{0\prime}\bar{c}_{R_0}\psi_{iT}\right]^2\\ &= T^{-2}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T\left(c_{R_0}'\lambda_i^0\lambda_i^{0\prime}\bar{c}_{R_0}\right)^2E_{\mathcal{D}}\left[\varepsilon_{it}^2\varepsilon_{is}^2\right]\\ &\quad+T^{-2}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^Tc_{R_0}'\lambda_i^0\lambda_j^{0\prime}\bar{c}_{R_0}\left[c_{R_0}'\lambda_i^0\lambda_j^{0\prime}\bar{c}_{R_0}+c_{R_0}'\lambda_j^0\lambda_i^{0\prime}\bar{c}_{R_0}\right]E_{\mathcal{D}}\left[\varepsilon_{it}^2\varepsilon_{jt}^2\right]\\ &= O_P\left(N+N^2T^{-1}\right).\end{aligned}$$

It follows that $T^{-1}\lambda^{0\prime}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\lambda^0-\lambda^{0\prime}\Psi_{NT}\lambda^0=O_P\left(N^{1/2}+NT^{-1/2}\right)$ by conditional Chebyshev inequality.

(iii) Noting that $\hat{\varepsilon}_i=M_{\hat{F}_{(1)}}(Y_i-X_{i(1)}\hat{\beta}_{(1)})$ and $Y_i-X_{i(1)}\hat{\beta}_{(1)}=(X_{i(1)}\beta_{(1)}^0+F_{(1)}^{*}\lambda_{i(1)}^{*}+\varepsilon_i)-X_{i(1)}\hat{\beta}_{(1)}=\hat{F}_{(1)}\lambda_{i(1)}^{*}+e_i$ where $e_i=\varepsilon_i+X_{i(1)}(\beta_{(1)}^0-\hat{\beta}_{(1)})+(F_{(1)}^{*}-\hat{F}_{(1)})\lambda_{i(1)}^{*}$, we have

$$\hat{\varepsilon}_i-\varepsilon_i=M_{\hat{F}_{(1)}}e_i-\varepsilon_i=-P_{\hat{F}_{(1)}}\varepsilon_i+M_{\hat{F}_{(1)}}\left[X_{i(1)}(\beta_{(1)}^0-\hat{\beta}_{(1)})+(F_{(1)}^{*}-\hat{F}_{(1)})\lambda_{i(1)}^{*}\right].$$

It follows that

$$
\begin{aligned}
\hat{\psi}_{iT}-\psi_{iT} &= T^{-1}\left(\hat{\varepsilon}_i'\hat{\varepsilon}_i-\varepsilon_i'\varepsilon_i\right)+T^{-1}\sum_{t=1}^{T}\left[\varepsilon_{it}^2-E_{\mathcal{D}}(\varepsilon_{it}^2)\right] \\
&= T^{-1}\left(e_i'M_{\hat{F}_{(1)}}e_i-\varepsilon_i'\varepsilon_i\right)+T^{-1}\sum_{t=1}^{T}\left[\varepsilon_{it}^2-E_{\mathcal{D}}(\varepsilon_{it}^2)\right] \\
&= -T^{-1}\varepsilon_i'P_{\hat{F}_{(1)}}\varepsilon_i+T^{-1}(\beta_{(1)}^0-\hat{\beta}_{(1)})'X_{i(1)}'M_{\hat{F}_{(1)}}X_{i(1)}(\beta_{(1)}^0-\hat{\beta}_{(1)}) \\
&\quad +T^{-1}\lambda_{i(1)}^{*\prime}(F_{(1)}^*-\hat{F}_{(1)})M_{\hat{F}_{(1)}}(F_{(1)}^*-\hat{F}_{(1)})\lambda_{i(1)}^*+2T^{-1}\varepsilon_i'M_{\hat{F}_{(1)}}X_{i(1)}(\beta_{(1)}^0-\hat{\beta}_{(1)}) \\
&\quad +2T^{-1}\varepsilon_i'M_{\hat{F}_{(1)}}(F_{(1)}^*-\hat{F}_{(1)})\lambda_{i(1)}^*+2(\beta_{(1)}^0-\hat{\beta}_{(1)})'T^{-1}X_{i(1)}'M_{\hat{F}_{(1)}}(F_{(1)}^*-\hat{F}_{(1)})\lambda_{i(1)}^* \\
&\quad +T^{-1}\sum_{t=1}^{T}\left[\varepsilon_{it}^2-E_{\mathcal{D}}(\varepsilon_{it}^2)\right] \\
&\equiv -D\psi_{iT,1}+D\psi_{iT,2}+D\psi_{iT,3}+2D\psi_{iT,4}+D\psi_{iT,5}+2D\psi_{iT,6}+D\psi_{iT,7},\ \text{say.}
\end{aligned}
$$

We bound each term in the last expression in order. First, by Lemmas A.2(vi) and D.2(iii),

$$
\begin{aligned}
\max_{1\le i\le N}\left|D\psi_{iT,1}\right| &= \max_{1\le i\le N}\left|T^{-1}\varepsilon_i'\left(P_{\hat{F}_{(1)}}-P_{F_{(1)}^*}\right)\varepsilon_i+T^{-1}\varepsilon_i'P_{F_{(1)}^*}\varepsilon_i\right| \\
&\le \left\|P_{\hat{F}_{(1)}}-P_{F_{(1)}^*}\right\|\max_{1\le i\le N}T^{-1}\left\|\varepsilon_i\right\|^2+\mu_{\max}\left(T^{-1}F^{0\prime}F^0\right)\max_{1\le i\le N}T^{-2}\left\|F^{0\prime}\varepsilon_i\right\|^2 \\
&= O_P\left(\delta_{NT}^{-1}\right)O_P(1)+O_P(1)\,O_P\left(\alpha_{NT}^2\right)=O_P\left(\delta_{NT}^{-1}\right).
\end{aligned}
$$

By Theorem 3.3 and Lemma D.2(v),

$$
\max_{1\le i\le N}\left|D\psi_{iT,2}\right|\le\left\|\beta_{(1)}^0-\hat{\beta}_{(1)}\right\|^2\max_{1\le i\le N}T^{-1}\left\|X_{i(1)}\right\|^2=O_P\left(K_0\delta_{NT}^{-4}\right)O_P\left(K_0\right)=O_P\left(K_0^2\delta_{NT}^{-4}\right).
$$

By Lemmas A.2(i) and D.2(vi),

$$
\max_{1\le i\le N}\left|D\psi_{iT,3}\right|\le T^{-1}\left\|F_{(1)}^*-\hat{F}_{(1)}\right\|^2\max_{1\le i\le N}\left\|\lambda_{i(1)}^*\right\|^2=O_P\left(\delta_{NT}^{-2}\right)o_P\left(N^{1/(4+2\sigma)}\right)=o_P\left(\delta_{NT}^{-2}N^{1/(4+2\sigma)}\right).
$$

Using $M_{\hat{F}_{(1)}}=I_T-P_{F^0}-(P_{\hat{F}_{(1)}}-P_{F^0})$ and $P_{F^0}=P_{F_{(1)}^*}$, we have that by Lemmas A.2(vi) and D.2(ii)-(v) and Assumption B.1(i),

$$
\begin{aligned}
\max_{1\le i\le N}\left|D\psi_{iT,4}\right| &\le \max_{1\le i\le N}\left|T^{-1}\varepsilon_i'X_{i(1)}(\beta_{(1)}^0-\hat{\beta}_{(1)})\right|+\max_{1\le i\le N}\left|T^{-1}\varepsilon_i'P_{F^0}X_{i(1)}(\beta_{(1)}^0-\hat{\beta}_{(1)})\right| \\
&\quad +\max_{1\le i\le N}\left|T^{-1}\varepsilon_i'(P_{\hat{F}_{(1)}}-P_{F_{(1)}^*})X_{i(1)}(\beta_{(1)}^0-\hat{\beta}_{(1)})\right| \\
&\le \left\|\beta_{(1)}^0-\hat{\beta}_{(1)}\right\|\left\{\max_{1\le i\le N}\left|T^{-1}\varepsilon_i'X_{i(1)}\right|\right. \\
&\quad +\left[\mu_{\min}\left(T^{-1}F^{0\prime}F^0\right)\right]^{-1}\max_{1\le i\le N}T^{-1}\left\|\varepsilon_i'F^0\right\|T^{-1/2}\left\|F^0\right\|\max_{1\le i\le N}T^{-1/2}\left\|X_{i(1)}\right\| \\
&\quad \left.+\left\|P_{\hat{F}_{(1)}}-P_{F_{(1)}^*}\right\|\max_{1\le i\le N}T^{-1/2}\left\|\varepsilon_i\right\|\max_{1\le i\le N}T^{-1/2}\left\|X_{i(1)}\right\|\right\} \\
&= O_P(K_0^{1/2}\delta_{NT}^{-2}) \\
&\quad \times\left\{O_P(K_0^{1/2}\alpha_{NT})+O_P(1)\,O_P(\alpha_{NT})\,O_P(1)\,O_P(K_0^{1/2})+O_P\left(\delta_{NT}^{-1}\right)O_P(1)\,O_P(K_0^{1/2})\right\} \\
&= o_P\left(\delta_{NT}^{-1}\right).
\end{aligned}
$$

Using $M_{\hat{F}_{(1)}} = I_T - P_{\hat{F}_{(1)}} = I_T - P_{F^0} + (P_{F^0} - P_{\hat{F}_{(1)}})$, the fact that $P_{F^0} = P_{F^*_{(1)}}$, and $P_{F^0} = F^0 \left(F^{0\prime}F^0\right)^{-1} F^{0\prime}$, we have that by Lemmas A.2(i) and (vi) and D.2(ii)-(v) and Assumption B.1(i),

$$\begin{aligned}
\max_{1\le i\le N} \left|D\psi_{iT,5}\right| &\le \max_{1\le i\le N} \left|T^{-1}\varepsilon_i'(F^*_{(1)} - \hat{F}_{(1)})\lambda^*_{i(1)}\right| + \max_{1\le i\le N} T^{-1}\left|\varepsilon_i' P_{F^0}(F^*_{(1)} - \hat{F}_{(1)})\lambda^*_{i(1)}\right| \\
&\quad + \max_{1\le i\le N} T^{-1}\left|\varepsilon_i'(P_{\hat{F}_{(1)}} - P_{F^*_{(1)}})(F^*_{(1)} - \hat{F}_{(1)})\lambda^*_{i(1)}\right| \\
&\le \max_{1\le i\le N} T^{-1/2}\left\|\varepsilon_i\right\| T^{-1/2}\left\|F^*_{(1)} - \hat{F}_{(1)}\right\| \left\|\lambda^*_{i(1)}\right\| \\
&\quad + \left[\mu_{\min}\left(T^{-1}F^{0\prime}F^0\right)\right]^{-1} \max_{1\le i\le N} T^{-1}\left\|\varepsilon_i' F^0\right\| T^{-1}\left\|F^{0\prime}(F^*_{(1)} - \hat{F}_{(1)})\right\| \max_{1\le i\le N}\left\|\lambda^*_{i(1)}\right\| \\
&\quad + \max_{1\le i\le N} T^{-1/2}\left\|\varepsilon_i\right\| \left\|P_{\hat{F}_{(1)}} - P_{F^*_{(1)}}\right\| T^{-1/2}\left\|F^*_{(1)} - \hat{F}_{(1)}\right\| \max_{1\le i\le N}\left\|\lambda^*_{i(1)}\right\| \\
&= O_P(1)\, O_P\left(\delta_{NT}^{-1}\right) o_P\left(N^{1/(8+4\sigma)}\right) + O_P(1)\, O_P(\alpha_{NT})\, O_P(\delta_{NT}^{-2} + (NT/K)^{-1/2}) \\
&\quad \times o_P\left(N^{1/(8+4\sigma)}\right) + O_P(1)\, O_P\left(\delta_{NT}^{-1}\right) O_P\left(\delta_{NT}^{-1}\right) o_P\left(N^{1/(8+4\sigma)}\right) \\
&= o_P\left(\delta_{NT}^{-1} N^{1/(8+4\sigma)}\right).
\end{aligned}$$

By CS inequality, $\max_{1\le i\le N}\left|D\psi_{iT,6}\right| \le \left\{\max_{1\le i\le N}\left|D\psi_{iT,2}\right| \max_{1\le i\le N}\left|D\psi_{iT,3}\right|\right\}^{1/2} = o_P(K_0\delta_{NT}^{-3}N^{1/(8+4\sigma)})$. Lastly, $\max_{1\le i\le N}\left|D\psi_{iT,7}\right| = O_P(\alpha_{NT})$ by Lemma D.2(i). Consequently, (iii) follows.

(iv) Observe that

$$\begin{aligned}
& (NT)^{-1}\left(\hat{F}_{(1)}\hat{\lambda}_{(1)}'\hat{\Psi}_{NT}\hat{\lambda}_{(1)}\hat{F}_{(1)}' - F^0\lambda^{0\prime}\Psi_{NT}\lambda^0 F^{0\prime}\right) \\
=\ & (NT)^{-1}\left(\hat{F}_{(1)}\hat{\lambda}_{(1)}' - F^0\lambda^{0\prime}\right)\Psi_{NT}\lambda^0 F^{0\prime} + (NT)^{-1}F^0\lambda^{0\prime}\Psi_{NT}\left(\hat{F}_{(1)}\hat{\lambda}_{(1)}' - F^0\lambda^{0\prime}\right) \\
& + (NT)^{-1}\left(\hat{F}_{(1)}\hat{\lambda}_{(1)}' - F^0\lambda^{0\prime}\right)\Psi_{NT}\left(\hat{F}_{(1)}\hat{\lambda}_{(1)}' - F^0\lambda^{0\prime}\right) + (NT)^{-1}\hat{F}_{(1)}\hat{\lambda}_{(1)}'\left(\hat{\Psi}_{NT} - \Psi_{NT}\right)\hat{\lambda}_{(1)}\hat{F}_{(1)}' \\
\equiv\ & I + II + III + IV, \text{ say.}
\end{aligned}$$

By the facts that

$$\begin{aligned}
\left\|\hat{F}_{(1)}\hat{\lambda}_{(1)}' - F^0\lambda^{0\prime}\right\| &\le \left\|(\hat{F}_{(1)} - F^*_{(1)})\lambda^{*\prime}_{(1)}\right\| + \left\|\hat{F}_{(1)}(\hat{\lambda}_{(1)} - \lambda^*_{(1)})'\right\| \\
&\le \left\|\hat{F}_{(1)} - F^*_{(1)}\right\| \left\|\lambda^*_{(1)}\right\| + \left\|\hat{F}_{(1)}\right\| \left\|\hat{\lambda}_{(1)} - \lambda^*_{(1)}\right\| \\
&= O_P\left(N^{1/2}T^{1/2}\delta_{NT}^{-1}\right) + O_P\left(N^{1/2}T^{1/2}\delta_{NT}^{-1}\right)
\end{aligned}$$

by Lemma A.2(i) and Theorem 3.1, $\|A\| \le$rank$(A)\,\|A\|_{\text{sp}}$, and that $\|A\|_{\text{sp}} \le \|A\|$, we have

$$\begin{aligned}
\|I\| &\le R_0\,(NT)^{-1}\max_{1\le i\le N}\psi_{iT}\left\|\hat{F}_{(1)}\hat{\lambda}' - F^0\lambda^{0\prime}\right\| \left\|\lambda^0 F^{0\prime}\right\| \\
&= (NT)^{-1} O_P(1)\, O_P\left(N^{1/2}T^{1/2}\delta_{NT}^{-1}\right) O_P\left(N^{1/2}T^{1/2}\right) = O_P\left(\delta_{NT}^{-1}\right).
\end{aligned}$$

Similarly, $\|II\| = \|I\| = O_P\left(\delta_{NT}^{-1}\right)$ and $\|III\| = O_P\left(\delta_{NT}^{-2}\right)$. In addition, $\|IV\| \le R_0\,(NT)^{-1}\,||\hat{F}_{(1)}\hat{\lambda}'||^2$ $\max_{1\le i\le N}|\hat{\psi}_{iT} - \psi_{iT}| = o_P(\delta_{NT}^{-1}N^{1/(8+4\sigma)})$ by (iii). Hence (vi) follows.

(v) Observe that $(NT)^{-1}\,(\hat{\lambda}'\hat{\Psi}_{NT}\mathbf{X}_k\hat{F}_{(1)} - H^+\lambda^{0\prime}\Psi_{NT}\mathbf{X}_k F^*_{(1)}) = (NT)^{-1}\,(\hat{\lambda} - \lambda^0 H^{+\prime})'\Psi_{NT}\mathbf{X}_k F^*_{(1)} + (NT)^{-1}\,\hat{\lambda}'\Psi_{NT}\mathbf{X}_k(\hat{F}_{(1)} - F^*_{(1)}) + (NT)^{-1}\,\hat{\lambda}'\,(\hat{\Psi}_{NT} - \Psi_{NT})\mathbf{X}_k\hat{F}_{(1)} \equiv J_1 + J_2 + J_3$, say. By Theorem 3.1

and Lemmas A.2(i) and D.3(iii), we can show that $J_s = O_P(\delta_{NT}^{-1})$ for $s = 1,2,3$. For example,

$$\begin{aligned}\|J_1\| &\leq R_0 (NT)^{-1} \max_{1\leq i\leq \leq N} \psi_{iT} \left\|\hat{\lambda} - \lambda^0 H^{+\prime}\right\| \|\mathbf{X}_k\| \left\|F^*_{(1)}\right\| \\ &= (NT)^{-1} O_P(1) O_P(N^{1/2}\delta_{NT}^{-1}) O_P(N^{1/2}T^{1/2}) O_P(T^{1/2}) = O_P(\delta_{NT}^{-1}).\end{aligned}$$

It follows that $(NT)^{-1} (\hat{\lambda}'\hat{\Psi}_{NT}\mathbf{X}_k\hat{F}_{(1)} - H^+\lambda^{0\prime}\Psi_{NT}\mathbf{X}_k F^*_{(1)}) = O_P(\delta_{NT}^{-1})$.

(vi) Let $ED \equiv N^{-3}E_{\mathcal{D}} \left\|\lambda^{0\prime}\left(T^{-1}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' - \Psi_{NT}\right)\mathbf{X}_k\right\|^2$. Note that

$$\begin{aligned}ED &= N^{-3}E_{\mathcal{D}}\mathrm{tr}\left\{\lambda^{0\prime}\left(T^{-1}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' - \Psi_{NT}\right)\mathbf{X}_k\mathbf{X}_k'\left(T^{-1}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' - \Psi_{NT}\right)\lambda^0\right\} \\ &= N^{-3}T^{-2}\sum_{1\leq i,j,l,m\leq N}\sum_{1\leq t,s,r\leq T}\zeta_{ijlm,tsr}\lambda_m^{0\prime}\lambda_i^0,\end{aligned}$$

where $\zeta_{ijlm,tsr} \equiv E_{\mathcal{D}}\left\{\left[\varepsilon_{it}\varepsilon_{jt} - E_{\mathcal{D}}\left(\varepsilon_{it}\varepsilon_{jt}\right)\right]\left[\varepsilon_{lr}\varepsilon_{mr} - E_{\mathcal{D}}\left(\varepsilon_{lr}\varepsilon_{mr}\right)\right]X_{js,k}X_{ls,k}\right\}$. We consider four cases for the individual indices $\{i,j,l,m\}$: (a) $\#\{i,j,l,m\} = 4$, (b) $\#\{i,j,l,m\} = 3$, (c) $\#\{i,j,l,m\} = 2$, and (d) $\#\{i,j,l,m\} = 1$. We use ED_a, ED_b, ED_c, and ED_d to denote ED when the individual indices in the summation are restricted to cases (a), (b), (c), and (d), respectively. In case (a), we can readily verify that $\zeta_{ijlm,tsr} = 0$ as $\zeta_{ijlm,tsr} = E_{\mathcal{D}}[\varepsilon_{it}]E_{\mathcal{D}}[\varepsilon_{jt}X_{js,k}]E_{\mathcal{D}}[\varepsilon_{lr}X_{ls,k}]E_{\mathcal{D}}[\varepsilon_{mr}] = 0$ by Assumption B.2(ii)-(iii) when $\#\{i,j,l,m\} = 4$. In case (b), wlog we consider three subcases: $(b1)$ $i = j$, $(b2)$ $i = l$, and $(b3)$ $i = m$ as the other cases can be analyzed analogously, and write the corresponding summations as ED_{b1}, ED_{b2}, and ED_{b3}, respectively. In subcase $(b1)$, $\zeta_{ijlm,tsr} = E_{\mathcal{D}}\{[\varepsilon_{it}^2 - E_{\mathcal{D}}\left(\varepsilon_{it}^2\right)]X_{is,k}\}E_{\mathcal{D}}(\varepsilon_{lr}X_{ls,k})E_{\mathcal{D}}(\varepsilon_{mr}) = 0$ by Assumption B.2(ii)-(iii) and thus $ED_{b1} = 0$. In subcase $(b2)$ $\zeta_{ijlm,tsr} = E_{\mathcal{D}}(\varepsilon_{it}\varepsilon_{ir}X_{is,k})E_{\mathcal{D}}(\varepsilon_{jt}X_{js,k})$ $E_{\mathcal{D}}(\varepsilon_{mr}) = 0$ by Assumption B.2(ii)-(iii) and thus $ED_{b2} = 0$. In subcase $(b3)$, we have $\zeta_{ijlm,tsr} = E_{\mathcal{D}}(\varepsilon_{it}\varepsilon_{ir})E_{\mathcal{D}}(\varepsilon_{jt}X_{js,k})E_{\mathcal{D}}(\varepsilon_{lr}X_{ls,k})$ by Assumption B.2(ii). In view of the facts that $E_{\mathcal{D}}(\varepsilon_{it}\varepsilon_{ir}) = 0$ if $t \neq r$, $E_{\mathcal{D}}(\varepsilon_{jt}X_{js,k}) = 0$ if $t \geq s$, and that $E_{\mathcal{D}}(\varepsilon_{lr}X_{ls,k}) = 0$ if $r \geq s$ by Assumption B.2(iii), we have

$$\begin{aligned}ED_{b3} &= N^{-3}T^{-2}\sum_{1\leq i\neq j\neq l\leq N}\sum_{1\leq t,s,r\leq T}E_{\mathcal{D}}(\varepsilon_{it}\varepsilon_{ir})E_{\mathcal{D}}(\varepsilon_{jt}X_{js,k})E_{\mathcal{D}}(\varepsilon_{lr}X_{ls,k})\left\|\lambda_i^0\right\|^2 \\ &= N^{-3}T^{-2}\sum_{1\leq i\neq j\neq l\leq N}\sum_{1\leq t<s\leq T}E_{\mathcal{D}}(\varepsilon_{it}^2)E_{\mathcal{D}}(\varepsilon_{jt}X_{js,k})E_{\mathcal{D}}(\varepsilon_{lt}X_{ls,k})\left\|\lambda_i^0\right\|^2.\end{aligned}$$

By the fact that $E_{\mathcal{D}}(\varepsilon_{jt}) = 0$ and Davydov inequality for conditional strong mixing processes (e.g., Su and Chen (2013, Lemma A.3)), we have $|E_{\mathcal{D}}(\varepsilon_{jt}X_{js,k})| \leq 8\left\|\varepsilon_{jt}\right\|_{8+4\sigma,\mathcal{D}}\left\|X_{js,k}\right\|_{8+4\sigma,\mathcal{D}}\alpha_{\mathcal{D}}\left(t-s\right)^{(3+2\sigma)/(4+2\sigma)}$ for any $t < s$ and any j. It follows that

$$\begin{aligned}|ED_{b3}| &\leq 64N^{-1}T^{-2}\sum_{i=1}^N\left\|\lambda_i^0\right\|^2\sum_{1\leq t<s\leq T}E_{\mathcal{D}}(\varepsilon_{it}^2)\left\{N^{-1}\sum_{j=1}^N\left\|\varepsilon_{jt}\right\|_{8+4\sigma,\mathcal{D}}\left\|X_{js,k}\right\|_{8+4\sigma,\mathcal{D}}\right\}^2\alpha_{\mathcal{D}}\left(s-t\right)^{(3+2\sigma)/(2+\sigma)} \\ &\leq 64c_{1NT}\left\{N^{-1}T^{-2}\sum_{i=1}^N\sum_{t=1}^T E_{\mathcal{D}}(\varepsilon_{it}^2)\left\|\lambda_i^0\right\|^2\right\}\sum_{\tau=1}^{\infty}\alpha_{\mathcal{D}}\left(\tau\right)^{(3+2\sigma)/(2+2\sigma)} \\ &= O_P(1)\,O_P\left(T^{-1}\right)O_P(1) = O_P\left(T^{-1}\right),\end{aligned}$$

where we use the facts that $c_{1NT} \equiv \max_{1\leq s,t\leq T} N^{-1}\sum_{j=1}^{N} \|\varepsilon_{jt}\|_{8+4\sigma,\mathcal{D}} \|X_{js,k}\|_{8+4\sigma,\mathcal{D}} = O_P(1)$ by Assumption B.1(v) and Cauchy-Schwarz inequality and that $\sum_{\tau=1}^{\infty} \alpha_{\mathcal{D}}(\tau)^{(3+2\sigma)/(2+2\sigma)} < \infty$ by Assumption B.2(i). It follows that $ED_b = O_P(T^{-1})$.

Now we consider case (c). We consider three subcases $(c1)$ $i=j\neq l=m$, $(c2)$ $i=l\neq j=m$, and $(c3)$ $i=m\neq j=l$, and use ED_{c1}, ED_{c2} and ED_{c3} to denote ED when the individual indices in its summation are restricted to these three subcases respectively. By Davydov inequality,

$$\begin{aligned}
|ED_{c1}| &= N^{-3}T^{-2}\left|\sum_{1\leq i\neq l\leq N}\sum_{1\leq t,s,r\leq T} E_{\mathcal{D}}\left\{[\varepsilon_{it}^2 - E_{\mathcal{D}}(\varepsilon_{it}^2)][\varepsilon_{lr}^2 - E_{\mathcal{D}}(\varepsilon_{lr}^2)]X_{is,k}X_{ls,k}\right\}\lambda_l^{0\prime}\lambda_i^0\right| \\
&\leq N^{-3}T^{-2}\sum_{1\leq i\neq l\leq N}\sum_{1\leq t\neq s\neq r\leq T}\left|E_{\mathcal{D}}\left\{[\varepsilon_{it}^2 - E_{\mathcal{D}}(\varepsilon_{it}^2)]X_{is,k}\right\}\right|\left|E_{\mathcal{D}}\left\{[\varepsilon_{lr}^2 - E_{\mathcal{D}}(\varepsilon_{lr}^2)]X_{ls,k}\right\}\right|\left\|\lambda_l^0\right\|\left\|\lambda_i^0\right\| \\
&\quad + O_P(N^{-1}) \\
&\leq 64N^{-3}T^{-2}\sum_{1\leq i\neq l\leq N}\sum_{1\leq t\neq s\neq r\leq T}\left\|\varepsilon_{it}^2\right\|_{4+2\sigma,\mathcal{D}}\|X_{is,k}\|_{4+2\sigma,\mathcal{D}}\left\|\varepsilon_{lr}^2\right\|_{4+2\sigma,\mathcal{D}}\|X_{ls,k}\|_{4+2\sigma,\mathcal{D}}\left\|\lambda_l^0\right\|\left\|\lambda_i^0\right\| \\
&\quad \times\alpha_{\mathcal{D}}(|s-t|)^{(1+\sigma)/(2+\sigma)}\alpha_{\mathcal{D}}(|s-r|)^{(1+\sigma)/(2+\sigma)} + O_P(N^{-1}) \\
&\leq 8c_{2NT}\max_{i,t}\left\|\varepsilon_{it}^2\right\|_{4+2\sigma,\mathcal{D}} N^{-2}T^{-2}\sum_{s=1}^{T}\sum_{i=1}^{N}\left\|\lambda_i^0\right\|\|X_{is,k}\|_{4+2\sigma,\mathcal{D}}\left\{\sum_{\tau=1}^{\infty}\alpha_{\mathcal{D}}(\tau)^{(1+\sigma)/(2+\sigma)}\right\}^2 \\
&\quad + O_P(N^{-1}) \\
&= O_P(1)\,o_P\left((NT)^{1/(4+2\sigma)}\right)O_P(N^{-1}T^{-1}) + O_P(N^{-1}) = o_P\left((NT)^{-1+1/(4+2\sigma)}\right) + O_P(N^{-1}),
\end{aligned}$$

where we use the facts that $c_{2NT} \equiv \max_{1\leq r,s\leq T} N^{-1}\sum_{l=1}^{N}\left\|\varepsilon_{lr}^2\right\|_{4+2\sigma,\mathcal{D}}[\|X_{ls,k}\|_{4+2\sigma,\mathcal{D}}\left\|\lambda_l^0\right\|] = O_P(1)$ by Assumption B.1(v) and Cauchy-Schwarz inequality, that $\max_{i,t}\left\|\varepsilon_{it}^2\right\|_{4+2\sigma,\mathcal{D}} = o_P((NT)^{1/(4+2\sigma)})$ as $E[\left\|\varepsilon_{it}^2\right\|_{4+2\sigma,\mathcal{D}}]^{4+2\sigma} = E|\varepsilon_{it}|^{8+4\sigma} < \infty$ by Assumption B.1(iii), and that $\sum_{\tau=1}^{\infty}\alpha_{\mathcal{D}}(\tau)^{(1+\sigma)/(2+\sigma)} = O_P(1)$ by Assumption B.2(i). Next, noting that $E_{\mathcal{D}}(\varepsilon_{it}\varepsilon_{ir}) = 0$ if $t\neq r$ and $E_{\mathcal{D}}(\varepsilon_{it}\varepsilon_{ir}X_{is,k}) = 0$ of $t\neq r$ and $\max(t,r)\geq s$ by Assumption B.2(iii), we apply Davydov inequality to obtain

$$\begin{aligned}
|ED_{c2}| &= N^{-3}T^{-2}\left|\sum_{1\leq i\neq j\leq N}\sum_{1\leq t,r<s\leq T} E_{\mathcal{D}}(\varepsilon_{it}\varepsilon_{ir}X_{is,k})\,E_{\mathcal{D}}(\varepsilon_{jt}\varepsilon_{jr}X_{js,k})\,\lambda_j^{0\prime}\lambda_i^0\right| \\
&\leq N^{-3}T^{-2}\sum_{1\leq i\neq j\leq N}\sum_{1\leq t,r<s\leq T}|E_{\mathcal{D}}(\varepsilon_{it}\varepsilon_{ir}X_{is,k})|\,|E_{\mathcal{D}}(\varepsilon_{jt}\varepsilon_{jr}X_{js,k})|\left\|\lambda_j^0\right\|\left\|\lambda_i^0\right\| \\
&\leq 2N^{-3}T^{-2}\sum_{1\leq i\neq j\leq N}\sum_{1\leq t<r<s\leq T}|E_{\mathcal{D}}(\varepsilon_{it}\varepsilon_{ir}X_{is,k})|\,|E_{\mathcal{D}}(\varepsilon_{jt}\varepsilon_{jr}X_{js,k})|\left\|\lambda_j^0\right\|\left\|\lambda_i^0\right\| + O_P(N^{-1}) \\
&\leq 128N^{-3}T^{-2}\sum_{1\leq i\neq j\leq N}\sum_{1\leq t<r<s\leq T}\|\varepsilon_{it}\varepsilon_{ir}\|_{4+2\sigma,\mathcal{D}}\|X_{is,k}\|_{8+4\sigma,\mathcal{D}}\|\varepsilon_{jt}\|_{8+4\sigma,\mathcal{D}}\|\varepsilon_{jr}X_{js,k}\|_{4+2\sigma,\mathcal{D}} \\
&\quad \times\left\|\lambda_j^0\right\|\left\|\lambda_i^0\right\|\alpha_{\mathcal{D}}(s-r)^{(5+4\sigma)/(8+4\sigma)}\alpha_{\mathcal{D}}(r-t)^{(5+4\sigma)/(8+4\sigma)} + O_P(N^{-1}) \\
&\leq 128c_{3nT}\max_{t}\|\varepsilon_{it}\|_{8+4\sigma,\mathcal{D}}\max_{i,s}\|X_{is,k}\|_{8+4\sigma,\mathcal{D}}\left\{N^{-2}T^{-2}\sum_{i=1}^{N}\sum_{r=1}^{T}\|\varepsilon_{ir}\|_{8+4\sigma,\mathcal{D}}\left\|\lambda_i^0\right\|\right\} \\
&\quad \times\left\{\sum_{\tau=1}^{\infty}\alpha_{\mathcal{D}}(\tau)^{(5+4\sigma)/(8+4\sigma)}\right\}^2 + O_P(N^{-1})
\end{aligned}$$

$$
\begin{aligned}
&= O_P(1)\, o_P\left((NT)^{1/(8+4\sigma)}\right) o_P\left((NT)^{1/(8+4\sigma)}\right) O_P\left(N^{-1}T^{-1}\right) O_P(1) + O_P\left(N^{-1}\right) \\
&= o_P\left((NT)^{-1+1/(4+2\sigma)}\right) + O_P\left(N^{-1}\right),
\end{aligned}
$$

where $c_{3NT} = \max_{t,s,r} N^{-1}\sum_{j=1}^{N} \|\varepsilon_{jt}\|_{8+4\sigma,\mathcal{D}} \|\varepsilon_{jr}X_{js,k}\|_{4+2\sigma,\mathcal{D}} \left\|\lambda_j^0\right\| = O_P(1)$ by Assumption B.1(v) and CS inequality. In addition, noting that $E_{\mathcal{D}}(\varepsilon_{it}\varepsilon_{ir}\varepsilon_{jt}\varepsilon_{jr}X_{js,k}X_{js,k}) = E_{\mathcal{D}}(\varepsilon_{it}\varepsilon_{ir}) E_{\mathcal{D}}(\varepsilon_{jt}\varepsilon_{jr}\ X_{js,k}X_{js,k})$ and the last expression is zero if either $t \neq r$ or $\max(t,r) > s$ under Assumptions B.2(iii), we have $|ED_{c3}| = N^{-3}T^{-2}\sum_{1\leq i\neq j\leq N}\sum_{1\leq t<s\leq T} E_{\mathcal{D}}\left(\varepsilon_{it}^2\right) E_{\mathcal{D}}(\varepsilon_{jt}^2 X_{js,k}^2) \left\|\lambda_i^0\right\|^2 = O_P\left(N^{-1}\right)$. In sum, $ED_c = O_P\left(N^{-1}+T^{-1}\right) + o_P((NT)^{-1+1/(4+2\sigma)}) = O_P(N^{-1}+T^{-1})$.

For case (d), we can also consider the application of Davydov inequality when the time indices (t,s,r) are all distinct. Straightforward calculation shows that $ED_d = o_P\left(N^{-2}(NT)^{1/(4+2\sigma)}\right) = o_P\left(N^{-1}\right)$. (Note that one can obtain a rough bound for ED_d by $N^{-3}T^{-2}O_P\left(NT^3\right) = O_P\left(N^{-2}T\right)$ without the need to apply Davydov inequality.) Consequently, we have shown that $ED = O_P\left(N^{-1}+T^{-1}\right)$. ■

Let Φ_{NT} =diag$(\varphi_{1N},...,\varphi_{TN})$ and $\varphi_{tN} = N^{-1}\sum_{i=1}^{N} E_{\mathcal{D}}\left[\varepsilon_{it}^2\right]$. Recall $\hat{\Phi}_{NT}$ =diag$(\hat{\varphi}_{1N},...,\hat{\varphi}_{TN})$ and $\hat{\varphi}_{tN} = N^{-1}\sum_{i=1}^{N}\hat{\varepsilon}_{it}^2$.

Lemma D.4 *Suppose the conditions in Corollary 3.4 hold. Then*
(i) $F^{0\prime}\left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right)F^0 = O_P\left(T^{1/2} + TN^{-1/2}\right)$;
(ii) $\left\|\hat{\Phi}_{NT} - \Phi_{NT}\right\|_{sp} = \max_{1\leq t\leq T}|\hat{\varphi}_{tN} - \varphi_{tN}| = o_P\left(\delta_{NT}^{-1}T^{1/(8+4\sigma)}\right)$;
(iii) $T^{-3}E_{\mathcal{D}}\left\|\mathbf{X}_k\left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right)F^0\right\|^2 = O_P\left(N^{-1}\right) + o_P\left(T^{-(7+4\sigma)/(8+4\sigma)}\right)$ *for* $k = 1,...,K_0$.

Proof. (i) The proof is analogous to that of Lemma D.3(ii) and thus omitted.

(ii) The proof is analogous to that of Lemma D.3(iii) and thus omitted.

(iii) Note that

$$
\begin{aligned}
& T^{-3}E_{\mathcal{D}}\left\|\mathbf{X}_k\left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right)F^0\right\|^2 \\
= \;& T^{-3}\text{tr}\left[\mathbf{X}_k\left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right)F^0F^{0\prime}\left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right)\mathbf{X}_k'\right] \\
= \;& T^{-3}N^{-2}\sum_{1\leq i\neq j\leq N}\sum_{1\leq t,s,r,q\leq T} E_{\mathcal{D}}\left\{\left[\varepsilon_{jt}\varepsilon_{js} - E_{\mathcal{D}}(\varepsilon_{jt}\varepsilon_{js})\right]\left[\varepsilon_{jr}\varepsilon_{jq} - E_{\mathcal{D}}(\varepsilon_{jr}\varepsilon_{jq})\right]\right\} E_{\mathcal{D}}\left[X_{it,k}X_{iq,k}\right]F_s^{0\prime}F_r^0 \\
& + T^{-3}N^{-2}\sum_{i=1}^{N}\sum_{1\leq t,s,r,q\leq T} E_{\mathcal{D}}\left\{\left[\varepsilon_{it}\varepsilon_{is} - E_{\mathcal{D}}(\varepsilon_{it}\varepsilon_{is})\right]\left[\varepsilon_{ir}\varepsilon_{iq} - E_{\mathcal{D}}(\varepsilon_{ir}\varepsilon_{iq})\right]X_{it,k}X_{iq,k}\right\}F_s^{0\prime}F_r^0 \\
\equiv \;& I + II, \text{ say,}
\end{aligned}
$$

where the second equality follows because $\varsigma_{ijl,rstq} \equiv E_{\mathcal{D}}\{\left[\varepsilon_{jt}\varepsilon_{js} - E_{\mathcal{D}}(\varepsilon_{jt}\varepsilon_{js})\right]\left[\varepsilon_{lr}\varepsilon_{lq} - E_{\mathcal{D}}(\varepsilon_{lr}\varepsilon_{lq})\right]X_{it,k}$ $X_{iq,k}\} = 0$ under Assumption B.2(ii)-(iii) if $\#\{i,j,k\} = 3$ or $j = i \neq l$ or $l = i \neq j$. To study I, we consider three cases for the time indices $\{t,s,r,q\}$ inside the summation: (a) $\#\{t,s,r,q\} = 4$, (b) $\#\{t,s,r,q\} = 3$, and (c) all other cases. We use I_a, I_b, and I_c to denote I when the time indices in the summation are restricted to cases (a), (b), and (c), respectively. Apparently, $I_a = 0$ and $I_c = O_P\left(T^{-1}\right)$

under Assumptions B.2(ii)-(iii) and B.1(i) and (iii). In case (b) noting that under Assumption B.2(iii) we always have

$$\begin{aligned} & E_{\mathcal{D}}\left\{\left[\varepsilon_{jt}\varepsilon_{js}-E_{\mathcal{D}}\left(\varepsilon_{jt}\varepsilon_{js}\right)\right]\left[\varepsilon_{jr}\varepsilon_{jq}-E_{\mathcal{D}}\left(\varepsilon_{jr}\varepsilon_{jq}\right)\right]\right\} \\ = \; & E_{\mathcal{D}}\left(\varepsilon_{jt}\varepsilon_{js}\varepsilon_{jr}\varepsilon_{jq}\right)-E_{\mathcal{D}}\left(\varepsilon_{jt}\varepsilon_{js}\right)E_{\mathcal{D}}\left(\varepsilon_{jr}\varepsilon_{jq}\right)=E_{\mathcal{D}}\left(\varepsilon_{jt}\varepsilon_{js}\varepsilon_{jr}\varepsilon_{jq}\right), \end{aligned}$$

wlog we can assume $r=q$ and then consider two subcases $(b1)$ $\max(t,s)>r$ and $(b2)$ $r>s>t$. Accordingly, we define I_{bs} as I_b but with the time indices restricted to subcase (bs) for $s=1,2$. Noting that $E_{\mathcal{D}}\left(\varepsilon_{jt}\varepsilon_{js}\varepsilon_{jr}\varepsilon_{jq}\right)=E_{\mathcal{D}}\left(\varepsilon_{jt}\varepsilon_{js}\varepsilon_{jr}^2\right)=0$ in subcase $(b1)$, $I_{b1}=0$. For subcase $(b2)$ we apply the Davydov inequality for conditional strong mixing processes to obtain

$$\left|E_{\mathcal{D}}\left(\varepsilon_{jt}\varepsilon_{js}\varepsilon_{jr}^2\right)\right|\leq 8\left\|\varepsilon_{jt}\varepsilon_{js}\right\|_{4+2\sigma,\mathcal{D}}\left\|\varepsilon_{jr}^2\right\|_{4+2\sigma,\mathcal{D}}\alpha_{\mathcal{D}}\left(r-s\right)^{(1+\sigma)/(2+\sigma)}.$$

Consequently, we have

$$\begin{aligned} |I_{b2}| &= T^{-3}N^{-2}\left|\sum_{1\leq i\neq j\leq N}\sum_{1\leq t<s<r\leq T}E_{\mathcal{D}}\left(\varepsilon_{jt}\varepsilon_{js}\varepsilon_{jr}^2\right)E_{\mathcal{D}}\left[X_{it,k}X_{ir,k}\right]F_s^{0\prime}F_r^0\right| \\ &\leq 8T^{-3}N^{-2}\sum_{1\leq i\neq j\leq N}\sum_{1\leq t<s<r\leq T}\left\|\varepsilon_{jt}\varepsilon_{js}\right\|_{4+2\sigma,\mathcal{D}}\left\|\varepsilon_{jr}^2\right\|_{4+2\sigma,\mathcal{D}}\alpha_{\mathcal{D}}\left(r-s\right)^{(1+\sigma)/(2+\sigma)}\left|E_{\mathcal{D}}\left[X_{it,k}X_{ir,k}\right]F_s^{0\prime}F_r^0\right| \\ &\leq 8T^{-1}c_{4NT}\max_{1\leq s\leq T}\left\|F_s^0\right\|\left\{T^{-2}N^{-1}\sum_{i=1}^N\sum_{t=1}^T\left\|X_{it,k}\right\|_{2,\mathcal{D}}\sum_{r=1}^T\left\|X_{ir,k}\right\|_{2,\mathcal{D}}\left\|F_r^0\right\|\right\}\sum_{\tau=1}^{\infty}\alpha_{\mathcal{D}}\left(\tau\right)^{(1+\sigma)/(2+\sigma)} \\ &= T^{-1}o_P\left(T^{1/(8+4\sigma)}\right)O_P\left(1\right)O_P\left(1\right)=o_P\left(T^{-(7+4\sigma)/(8+4\sigma)}\right), \end{aligned}$$

where we use Lemma D.5(i) below, the fact that $c_{4NT}\equiv\max_{1\leq t,s,r\leq T}N^{-1}\sum_{j=1}^N\left\|\varepsilon_{jt}\right\|_{8+4\sigma,\mathcal{D}}\left\|\varepsilon_{js}\right\|_{8+4\sigma,\mathcal{D}}$ $\times\left\|\varepsilon_{jr}^2\right\|_{4+2\sigma,\mathcal{D}}=O_P\left(1\right)$ by Assumption B.1(v) and Cauchy-Schwarz inequality, and the fact that

$$\begin{aligned} & T^{-2}N^{-1}\sum_{i=1}^N\sum_{t=1}^T\left\|X_{it,k}\right\|_{2,\mathcal{D}}\sum_{r=1}^T\left\|X_{ir,k}\right\|_{2,\mathcal{D}}\left\|F_r^0\right\| \\ \leq \; & \frac{1}{2}T^{-1}N^{-1}\sum_{i=1}^N\sum_{t=1}^T E_{\mathcal{D}}\left(X_{it,k}^2\right)+\frac{1}{2}T^{-1}N^{-1}\sum_{i=1}^N\sum_{r=1}^T E_{\mathcal{D}}\left(X_{ir,k}^2\right)\left\|F_r^0\right\|^2=O_P\left(1\right)+O_P\left(1\right)=O_P\left(1\right) \end{aligned}$$

by Cauchy-Schwarz and Markov inequalities. It follows that $I=o_P(T^{-(7+4\sigma)/(8+4\sigma)})$.

For II, we consider two cases: (a) $\#\left\{t,s,r,q\right\}=4$ and (b) $\#\left\{t,s,r,q\right\}\leq 3$, and write $II=II_a+II_b$ where II_l is defined as II but with the time indices restricted to case (l) for $l=a,b$. Apparently, $II_b=T^{-3}N^{-2}O_P\left(NT^2\right)=O_P\left(N^{-1}\right)$. For case (a), $E_{\mathcal{D}}\left\{\left[\varepsilon_{it}\varepsilon_{is}-E_{\mathcal{D}}\left(\varepsilon_{it}\varepsilon_{is}\right)\right]\left[\varepsilon_{ir}\varepsilon_{iq}-E_{\mathcal{D}}\left(\varepsilon_{ir}\varepsilon_{iq}\right)\right]X_{it,k}X_{iq,k}\right\}$ $=E_{\mathcal{D}}\left[\varepsilon_{it}\varepsilon_{is}\varepsilon_{ir}\varepsilon_{iq}X_{it,k}X_{iq,k}\right]=0$ and thus $II_a=0$. It follows that $II=O_P\left(N^{-1}\right)$. ■

Lemma D.5 *Suppose the conditions in Corollary 3.4 hold. Let* $\varepsilon_{\cdot t}\equiv\left(\varepsilon_{1t},...,\varepsilon_{Nt}\right)'$ *and* $\hat{\varepsilon}_{\cdot t}\equiv\left(\hat{\varepsilon}_{1t},...,\hat{\varepsilon}_{Nt}\right)'$. *Let* $\bar{X}_{\cdot s,k}$ *denote the sth column of the* $N\times T$ *matrix* $\bar{\mathbf{X}}_k$. *Then*

(i) $\max_{1\leq t\leq T}\left\|F_t^0\right\|=o_P\left(T^{1/(8+4\sigma)}\right)$;

(ii) $\max_{1\leq t\leq T}N^{-1}\left\|\varepsilon_{\cdot t}\right\|^2=O_P\left(1\right)$;

(iii) $\max_{1\leq t\leq T} N^{-1}\left\|\lambda^{0\prime}\varepsilon_{\cdot t}\right\| = \max_{1\leq t\leq T}\left\|N^{-1}\sum_{i=1}^{N}\lambda_i^0\varepsilon_{it}\right\| = O_P\left(\alpha_{TN}\right);$
(iv) $\max_{1\leq t\leq T} N^{-1}\left\|X_{\cdot t}\right\|^2 = O_P\left(K\right);$
(v) $\max_{1\leq t\leq T-M}\max_{t<s\leq t+M} N^{-1}\left|\varepsilon_{\cdot t}'\bar{X}_{\cdot s,k} - E_{\mathcal{D}}\left(\varepsilon_{\cdot t}'\bar{X}_{\cdot s,k}\right)\right| = O_P\left(\alpha_{TN}\right);$
(vi) $\max_{1\leq t\leq T}\left\|\hat{F}_{t(1)} - F_{t(1)}^*\right\| = O_P\left(\eta_{1NT}\right);$
(vii) $\max_{1\leq t\leq T}\left\|\hat{F}_{t(1)}\right\| = O_P\left(T^{1/2}\delta_{NT}^{-1}\right) + o_P\left(T^{1/(8+4\sigma)}\right);$
(viii) $\max_{1\leq i\leq N}\max_{1\leq t\leq T}\left|\hat{\varepsilon}_{it} - \varepsilon_{it}\right| = o_P\left(1\right);$
(ix) $\max_{1\leq t\leq T} N^{-1}\left\|\hat{\varepsilon}_{\cdot t} - \varepsilon_{\cdot t}\right\|^2 = o_P\left(1\right);$
where $\eta_{1NT} = O_P\left(\delta_{NT}^{-1}\right) + o_P\left(\alpha_{TN} + N^{-1/2}T^{1/(8+4\sigma)}\right)$ *and* α_{TN} *is analogously defined as* α_{NT} *by interchanging* N *and* T.

Proof. (i) The proof is analogous to that of Lemma D.2(vi).

(ii) Note that $\max_{1\leq t\leq T} N^{-1}\left\|\varepsilon_{\cdot t}\right\|^2 \leq \max_{1\leq t\leq T}|N^{-1}\sum_{i=1}^{N}\left[\varepsilon_{it}^2 - E_{\mathcal{D}}\left(\varepsilon_{it}^2\right)\right]| + \max_{1\leq t\leq T}|N^{-1}\sum_{i=1}^{N} E_{\mathcal{D}}\left(\varepsilon_{it}^2\right)|$. Analogously to the proof of Lemma D.2(ii), we can show the first term is $O_P\left(\alpha_{TN}\right)$. The second term is $O_P\left(1\right)$ by Assumption B.1(iv). Thus (ii) follows.

(iii) The proof is analogous to that of Lemma D.2(ii).

(iv) The proof is analogous to that of (ii).

(v) The proof is analogous to that of Lemma D.2(ii).

(vi) Write $\hat{F}_{t(1)} - F_{t(1)}^* = \sum_{l=1}^{8} a_{lt(1)}$ where $a_{lt(1)}$ denotes the transpose of the tth row of $a_{l(1)}$ and recall $a_l = (a_{l(1)}, a_{l(2)})$, $l = 1, 2, ..., 8$, are defined in (C.2). By (i)-(iv), we can readily show that $\max_{1\leq t\leq T}||a_{1t(1)}|| = O_P(\delta_{NT}^{-1})$, $\max_{1\leq t\leq T}||a_{2t(1)}|| = o_P(N^{-1/2}T^{1/(8+4\sigma)})$, $\max_{1\leq t\leq T}||a_{3t(1)}|| = o_P\left(\alpha_{TN}\right)$, $\max_{1\leq t\leq T}||a_{4t(1)}|| = O_P((NT/K)^{-1})$, $\max_{1\leq t\leq T}||a_{5t(1)}|| = o_P((NT/K)^{-1/2}T^{1/(8+4\sigma)})$, $\max_{1\leq t\leq T}||a_{6t(1)}|| = O_P((NT/K)^{-1/2})$, $\max_{1\leq t\leq T}||a_{7t(1)}|| = o_P(K^{1/2}N^{-1}T^{-1/2}T^{1/(8+4\sigma)})$, and $\max_{1\leq t\leq T}||a_{8t(1)}|| = O_P(K^{1/2}\left(NT\right)^{-1}(N^{1/2} + T^{1/2}))$, where, e.g.,

$$\begin{aligned}\max_{1\leq t\leq T}\left\|a_{1t(1)}\right\| &= \max_{1\leq t\leq T}\left(NT\right)^{-1}\left\|\tilde{F}_{(1)}'\boldsymbol{\varepsilon}'\varepsilon_{\cdot t}\right\| \leq R_0\left(NT\right)^{-1}\left\|\tilde{F}_{(1)}\right\|\left\|\boldsymbol{\varepsilon}\right\|_{\text{sp}}\max_{1\leq t\leq T}\left\|\varepsilon_{\cdot t}\right\| \\ &= \left(NT\right)^{-1}O_P\left(T^{1/2}\right)O_P\left(N^{1/2}+T^{1/2}\right)O_P\left(N^{1/2}\right) = O_P\left(\delta_{NT}^{-1}\right).\end{aligned}$$

It follows that $\max_{1\leq t\leq T}\left\|\hat{F}_{t(1)} - F_{t(1)}^0\right\| = O_P\left(\eta_{1NT}\right)$.

(vii) By Lemmas A.2(i) and D.5(i), $\max_t||\hat{F}_{t(1)}|| \leq \max_t||\hat{F}_{t(1)} - F_t^0 H_{(1)}|| + \max_t||F_t^0 H_{(1)}|| = O_P(T^{1/2}\delta_{NT}^{-1}) + o_P(T^{1/(8+4\sigma)})$.

(viii) Noting that $\hat{\varepsilon}_i = M_{\hat{F}_{(1)}}(Y_i - X_{i(1)}\hat{\beta}_{(1)})$ and $Y_i - X_{i(1)}\hat{\beta}_{(1)} = (X_{i(1)}\beta_{(1)}^0 + F_{(1)}^*\lambda_{i(1)}^* + \varepsilon_i) - X_{i(1)}\hat{\beta}_{(1)} = \hat{F}_{(1)}\lambda_{i(1)}^* + e_i$ where $e_i = \varepsilon_i + X_{i(1)}(\beta_{(1)}^0 - \hat{\beta}_{(1)}) + (F_{(1)}^* - \hat{F}_{(1)})\lambda_{i(1)}^*$, we have

$$\hat{\varepsilon}_i - \varepsilon_i = M_{\hat{F}_{(1)}}e_i - \varepsilon_i = -P_{\hat{F}_{(1)}}\varepsilon_i + M_{\hat{F}_{(1)}}\left[X_{i(1)}(\beta_{(1)}^0 - \hat{\beta}_{(1)}) + (F_{(1)}^* - \hat{F}_{(1)})\lambda_{i(1)}^*\right].$$

It follows that

$$
\begin{aligned}
\hat{\varepsilon}_{it}-\varepsilon_{it} &= \hat{F}_{t(1)}'\left(\hat{F}_{(1)}'\hat{F}_{(1)}\right)^{-1}\hat{F}_{(1)}'\varepsilon_i+X_{it(1)}(\beta_{(1)}^0-\hat{\beta}_{(1)})+(F_{t(1)}^*-\hat{F}_{t(1)})'\lambda_{i(1)}^* \\
&\quad -\hat{F}_{t(1)}'\left(\hat{F}_{(1)}'\hat{F}_{(1)}\right)^{-1}\hat{F}_{(1)}'X_{i(1)}(\beta_{(1)}^0-\hat{\beta}_{(1)})-\hat{F}_{t(1)}'\left(\hat{F}_{(1)}'\hat{F}_{(1)}\right)^{-1}\hat{F}_{(1)}'(F_{(1)}^*-\hat{F}_{(1)})\lambda_{i(1)}^* \\
&\equiv \vartheta_{1it}+\vartheta_{2it}+\vartheta_{3it}-\vartheta_{4it}-\vartheta_{5it},\ \text{say.} \qquad \text{(D.4)}
\end{aligned}
$$

Noting that $\vartheta_{1it}=\hat{F}_{t(1)}'(\hat{F}_{(1)}'\hat{F}_{(1)})^{-1}(\hat{F}_{(1)}-F^0H_{(1)})'\varepsilon_i+\hat{F}_{t(1)}'(\hat{F}_{(1)}'\hat{F}_{(1)})^{-1}H_{(1)}'F^{0\prime}\varepsilon_i$, we have by Lemmas D.2(ii)-(iii) and D.5(vi)

$$
\begin{aligned}
\max_{i,t}|\vartheta_{1it}| &\leq T^{-1}\max_t\left\|\hat{F}_{t(1)}\right\|\left\|\left(T^{-1}\hat{F}_{(1)}'\hat{F}_{(1)}\right)^{-1}\right\|\left\|\hat{F}_{(1)}-F^0H_{(1)}\right\|\max_i\|\varepsilon_i\| \\
&\quad +T^{-1}\max_t\left\|\hat{F}_{t(1)}\right\|\left\|\left(T^{-1}\hat{F}_{(1)}'\hat{F}_{(1)}\right)^{-1}\right\|\|H_{(1)}\|\max_i\|F^0\varepsilon_i\| \\
&= T^{-1}\left[O_P\left(T^{1/2}\delta_{NT}^{-1}\right)+o_P\left(T^{1/(8+4\sigma)}\right)\right]O_P(1)\,O_P\left(T^{1/2}\delta_{NT}^{-1}\right)O_P\left(T^{1/2}\right) \\
&\quad +T^{-1}\left[O_P\left(T^{1/2}\delta_{NT}^{-1}\right)+o_P\left(T^{1/(8+4\sigma)}\right)\right]O_P(1)\,O_P(1)\,O_P(T\alpha_{NT}) \\
&= O_P\left(T^{1/2}\delta_{NT}^{-2}+T^{1/2}\delta_{NT}^{-1}\alpha_{NT}\right)+o_P\left(T^{1/(8+4\sigma)}\right)O_P\left(\delta_{NT}^{-1}+\alpha_{NT}\right)=o_P(1)\,.
\end{aligned}
$$

Noting that $\max_{i,t}\left\|X_{it(1)}\right\| = o_P\left((NT)^{1/(8+4\sigma)}\right)$ by Assumption B.1(iii) and Markov inequality, by Theorem 3.1 we have

$$
\max_{i,t}|\vartheta_{2it}|\leq\left\|\beta_{(1)}^0-\hat{\beta}_{(1)}\right\|\max_{i,t}\|X_{it(1)}\|=o_P\left(K_0^{1/2}\delta_{NT}^{-2}(NT)^{1/(8+4\sigma)}\right)=o_P(1)\,.
$$

By Lemmas A.2(i) and (iv), Lemmas D.2(v)-(vi), and Theorem 3.1, we have

$$
\begin{aligned}
\max_{i,t}|\vartheta_{3it}| &\leq \max_t\left\|F_{t(1)}^*-\hat{F}_{t(1)}\right\|\max_i\left\|\lambda_{i(1)}^*\right\| \\
&= \left[O_P\left(\delta_{NT}^{-1}\right)+o_P\left(\alpha_{TN}+N^{-1/2}T^{1/(8+4\sigma)}\right)\right]o_P\left(N^{1/(8+4\sigma)}\right)=o_P(1)\,, \\
\max_{i,t}|\vartheta_{4it}| &\leq \left\|\beta_{(1)}^0-\hat{\beta}_{(1)}\right\|\max_t\left\|\hat{F}_{t(1)}\right\|\left\|\left(\hat{F}_{(1)}'\hat{F}_{(1)}\right)^{-1}\hat{F}_{(1)}'\right\|\max_i\|X_{i(1)}\| \\
&= O_P\left(\delta_{NT}^{-2}\right)\left[O_P\left(T^{1/2}\delta_{NT}^{-1}\right)+o_P\left(T^{1/(8+4\sigma)}\right)\right]O_P\left(T^{-1/2}\right)O_P\left(T^{1/2}\right)=o_P(1)\,,\ \text{and} \\
\max_{i,t}|\vartheta_{5it}| &\leq \max_t\left\|\hat{F}_{t(1)}\right\|\left\|\left(\hat{F}_{(1)}'\hat{F}_{(1)}\right)^{-1}\right\|\left\|\hat{F}_{(1)}'(F_{(1)}^0-\hat{F}_{(1)})\right\|\max_i\left\|\lambda_{i(1)}^*\right\| \\
&= \left[O_P\left(T^{1/2}\delta_{NT}^{-1}\right)+o_P\left(T^{1/(8+4\sigma)}\right)\right]O_P\left(T^{-1}\right)O_P\left(T\delta_{NT}^{-2}\right)o_P\left(N^{1/(8+4\sigma)}\right)=o_P(1)\,.
\end{aligned}
$$

(ix) $\max_{1\leq t\leq T}N^{-1}\left\|\hat{\varepsilon}_{\cdot t}-\varepsilon_{\cdot t}\right\|^2=\max_{1\leq t\leq T}N^{-1}\sum_{i=1}^N\left(\hat{\varepsilon}_{it}-\varepsilon_{it}\right)^2\leq\max_{1\leq i\leq N}\max_{1\leq t\leq T}\left(\hat{\varepsilon}_{it}-\varepsilon_{it}\right)^2$ $=o_P(1)$ by (viii). ■

E Proof of Corollary 3.4

By the definitions of $\hat{\beta}_{(1)}^c$ and $\mathbb{B}_{NT}$, we have

$$
\mathbb{C}_{K_0}\sqrt{NT}(\hat{\beta}_{(1)}^c-\beta_{(1)}^0)=\mathbb{C}_{K_0}[\sqrt{NT}(\hat{\beta}_{(1)}-\beta_{(1)}^0)-\mathbb{B}_{NT}]-\mathbb{R}_{NT},
$$

where $\mathbb{R}_{NT}=\mathbb{C}_{K_0}\hat{D}_{\hat{F}_{(1)}}^{-1}[(\hat{\mathbb{B}}_{1NT}-\mathbb{B}_{1NT})-(\hat{\mathbb{B}}_{2NT}-\mathbb{B}_{2NT})-(\hat{\mathbb{B}}_{3NT}-\mathbb{B}_{3NT})-(\hat{\mathbb{B}}_{4NT}-\mathbb{B}_{4NT})]$. By the proof of Theorem 3.3, $\left\|\hat{D}_{\hat{F}_{(1)}}-D_{F^0}\right\|_{\rm sp}=o_P(1)$ and $\mathbb{C}_{K_0}[\sqrt{NT}(\hat{\beta}_{(1)}-\beta^0_{(1)})-\mathbb{B}_{NT}]\xrightarrow{d}N\left(0,\ \lim_{(N,T)\to\infty}\mathbb{C}_{K_0}\mathbb{V}_{NT}\mathbb{C}_{K_0}'\right)$. It suffices to show that $\hat{\mathbb{B}}_{lNT}-\mathbb{B}_{lNT}=\mathbf{o}_P(1)$ for $l=1,2,3,4$, and $\left\|\hat{\mathbb{V}}_{NT}-\mathbb{V}_{NT}\right\|_{\rm sp}=o_P(1)$.

First, we show (i) $\hat{\mathbb{B}}_{1NT}-\mathbb{B}_{1NT}=\mathbf{o}_P(1)$. Let $\bar{\mathbb{B}}_{1NT}=N^{-5/2}T^{-3/2}\sum_{i=1}^N X_{i(1)}'F^*_{(1)}[F^{*\prime}_{(1)}F^*_{(1)}]^{-1}\tilde{F}_{(1)}'F^0\lambda^{0\prime}\Psi_{NT}\lambda^0F^{0\prime}\tilde{F}_{(1)}\lambda^*_{(1)}$. We prove (i) by showing that (i1) $\bar{\mathbb{B}}_{1NT}-\mathbb{B}_{1NT}=\mathbf{o}_P(1)$ and (i2) $\hat{\mathbb{B}}_{1NT}-\bar{\mathbb{B}}_{1NT}=\mathbf{o}_P(1)$. To show (i1), let $c_{K_0}=(c_{1K_0},...,c_{K_0K_0})'$ be an arbitrary $K_0\times 1$ nonrandom vector such that $\|c_{K_0}\|=1$. By Lemma D.3(ii) and Assumptions A.3(i) and A.6(i), we have

$$
\begin{aligned}
&\left|c_{K_0}'\left(\mathbb{B}_{1NT}-\bar{\mathbb{B}}_{1NT}\right)\right|\\
=\;& N^{-5/2}T^{-3/2}\left|\sum_{k=1}^{K_0}c_{kK_0}\mathrm{tr}\left\{F^*_{(1)}(F^{*\prime}_{(1)}F^*_{(1)})^{-1}\tilde{F}_{(1)}'F^0\left[\lambda^{0\prime}(T^{-1}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'-\Psi_{NT})\lambda^0\right]F^{0\prime}\tilde{F}_{(1)}\lambda^{*\prime}_{(1)}\mathbf{X}_k\right\}\right|\\
\le\;& N^{-5/2}T^{-3/2}\left\|\lambda^{0\prime}(T^{-1}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'-\Psi_{NT})\lambda^0\right\|\left\|F^*_{(1)}(F^{*\prime}_{(1)}F^*_{(1)})^{-1}\right\|\left\|\tilde{F}_{(1)}'F^0\right\|^2||\lambda^*_{(1)}||\left\{\sum_{k=1}^{K_0}\|\mathbf{X}_k\|^2\right\}^{1/2}\\
=\;& N^{-5/2}T^{-3/2}O_P\left(N^{1/2}+NT^{-1/2}\right)O_P\left(T^{-1/2}\right)O_P\left(T^2\right)O_P\left(N^{1/2}\right)O_P\left((K_0NT)^{1/2}\right)\\
=\;& O_P\left(K_0^{1/2}(N^{-1}T^{1/2}+N^{-1/2}T^{-1/2})\right)=o_P(1).
\end{aligned}
$$

To show (i2), we make the following decomposition:

$$
\begin{aligned}
&\hat{\mathbb{B}}_{1NT}-\bar{\mathbb{B}}_{1NT}\\
=\;& N^{-5/2}T^{-3/2}\sum_{i=1}^N X_{i(1)}'[\hat{F}_{(1)}(\hat{F}_{(1)}'\hat{F}_{(1)})^{-1}-F^*_{(1)}(F^{*\prime}_{(1)}F^*_{(1)})^{-1}]\tilde{F}_{(1)}'F^0\lambda^{0\prime}\Psi_{NT}\lambda^0F^{0\prime}\tilde{F}_{(1)}\lambda^*_{i(1)}\\
&+N^{-5/2}T^{-3/2}\sum_{i=1}^N X_{i(1)}'\hat{F}_{(1)}(\hat{F}_{(1)}'\hat{F}_{(1)})^{-1}\tilde{F}_{(1)}'[\hat{F}_{(1)}\hat{\lambda}_{(1)}'\hat{\Psi}_{NT}\hat{\lambda}_{(1)}\hat{F}_{(1)}'-F^0\lambda^{0\prime}\Psi_{NT}\lambda^0F^{0\prime}]\tilde{F}_{(1)}\lambda^*_{i(1)}\\
&+N^{-5/2}T^{-3/2}\sum_{i=1}^N X_{i(1)}'\hat{F}_{(1)}(\hat{F}_{(1)}'\hat{F}_{(1)})^{-1}\tilde{F}_{(1)}'\hat{F}_{(1)}\hat{\lambda}_{(1)}'\hat{\Psi}_{NT}\hat{\lambda}_{(1)}\hat{F}_{(1)}'\tilde{F}_{(1)}(\hat{\lambda}_{i(1)}-\lambda^*_{i(1)})\\
\equiv\;& DB_{1NT,1}+DB_{1NT,2}+DB_{1NT,3},\ \text{say.}
\end{aligned}
$$

Let $\nu_{1NT}=\left\|\sum_{i=1}^N\lambda^*_{i(1)}c_{K_0}'X_{i(1)}\right\|_{\rm sp}$. Following the analysis of ς_{2NT} in the proof of Proposition B.1, we can readily show that $\nu_{1NT}=O_P(NT^{1/2})$. Then by Lemmas A.2(i) and (v), D.3(i) and (iv) and Assumption A.3(i), we have

$$
\begin{aligned}
\left|c_{K_0}'DB_{1NT,1}\right| &\le N^{-5/2}T^{-3/2}\left\|\left[\hat{F}_{(1)}\left(\hat{F}_{(1)}'\hat{F}_{(1)}\right)^{-1}-F^*_{(1)}\left(F^{*\prime}_{(1)}F^*_{(1)}\right)^{-1}\right]\left\|\tilde{F}_{(1)}'F^0\right\|^2\left\|\lambda^{0\prime}\Psi_{NT}\lambda^0\right\|\right\|\nu_{1NT}\\
&= N^{-5/2}T^{-3/2}O_P(T^{-1/2}\delta_{NT}^{-1})O_P\left(T^2\right)O_P(N)O_P(NT^{1/2})=O_P(N^{-1/2}T^{1/2}\delta_{NT}^{-1})=o_P(1),
\end{aligned}
$$

and

$$\begin{aligned}\left|c_{K_0}' DB_{1NT,2}\right| &\leq N^{-5/2}T^{-3/2}\left\|\hat{F}_{(1)}\left(\hat{F}_{(1)}'\hat{F}_{(1)}\right)^{-1}\right\|\left\|\hat{F}_{(1)}\hat{\lambda}_{(1)}'\hat{\Psi}_{NT}\hat{\lambda}_{(1)}\hat{F}_{(1)}' - F^0\lambda^{0\prime}\Psi_{NT}\lambda^0F^{0\prime}\right\|\left\|\tilde{F}_{(1)}\right\|^2\nu_{1NT} \\ &= N^{-5/2}T^{-3/2}O_P\left(T^{-1/2}\right)o_P\left(NT\delta_{NT}^{-1}N^{1/(8+4\sigma)}\right)O_P\left(T\right)O_P(NT^{1/2}) \\ &= o_P(N^{-1/2}T^{1/2}\delta_{NT}^{-1}N^{1/(8+4\sigma)}) = o_P\left(1\right).\end{aligned}$$

It follows that $\|DB_{1NT,l}\| = o_P(1)$ for $l = 1, 2$. In addition,

$$\begin{aligned}\|DB_{1NT,3}\| &\leq N^{-5/2}T^{-3/2}\left\|\hat{F}_{(1)}\left(\hat{F}_{(1)}'\hat{F}_{(1)}\right)^{-1}\right\|\left\|\hat{F}_{(1)}\hat{\lambda}_{(1)}'\hat{\Psi}_{NT}\hat{\lambda}_{(1)}\hat{F}_{(1)}'\right\|\left\|\tilde{F}_{(1)}\right\|^2\sum_{i=1}^N\left\|X_{i(1)}\right\|\left\|\hat{\lambda}_{i(1)} - \lambda_{i(1)}^*\right\| \\ &\leq N^{-5/2}T^{-3/2}\left\|\hat{F}_{(1)}\left(\hat{F}_{(1)}'\hat{F}_{(1)}\right)^{-1}\right\|\left\|\hat{F}_{(1)}\hat{\lambda}_{(1)}'\hat{\Psi}_{NT}\hat{\lambda}_{(1)}\hat{F}_{(1)}'\right\|\left\|\tilde{F}_{(1)}\right\|^2\nu_{2NT}^{1/2}\left\|\hat{\lambda}_{(1)} - \lambda_{(1)}^*\right\| \\ &= N^{-5/2}T^{-3/2}O_P\left(T^{-1/2}\right)O_P\left(NT\right)O_P\left(T\right)O_P(K_0^{1/2}N^{1/2}T^{1/2})O_P\left(N^{1/2}T^{-1/2}\right) \\ &= O_P(K_0^{1/2}N^{-1/2}) = o_P\left(1\right) \text{ under Assumption A.3(i),}\end{aligned}$$

where $\nu_{2NT} = \sum_{i=1}^N\left\|X_{i(1)}\right\|^2 = O_P(K_0NT)$. It follows that $\hat{\mathbb{B}}_{1NT} - \bar{\mathbb{B}}_{1NT} = \mathbf{o}_P(1)$ and $\hat{\mathbb{B}}_{1NT} - \mathbb{B}_{1NT} = \mathbf{o}_P(1)$.

Second, we prove (ii) $\hat{\mathbb{B}}_{2NT} - \mathbb{B}_{2NT} = \mathbf{o}_P(1)$. Let $\bar{\mathbb{B}}_{2NT} \equiv N^{-1/2}T^{-3/2}\sum_{i=1}^N X_{i(1)}'M_{F^0}\Phi_{NT}\tilde{F}_{(1)}\lambda_{i(1)}^*$. We prove (ii) by showing that (ii1) $\mathbb{B}_{2NT} - \bar{\mathbb{B}}_{2NT} = \mathbf{o}_P(1)$ and (ii2) $\hat{\mathbb{B}}_{2NT} - \bar{\mathbb{B}}_{2NT} = \mathbf{o}_P(1)$. Note that

$$\begin{aligned}\mathbb{B}_{2NT} - \bar{\mathbb{B}}_{2NT} &= N^{-1/2}T^{-3/2}\sum_{i=1}^N X_{i(1)}'\left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right)F^0H_{(1)}\lambda_{i(1)}^* \\ &\quad -N^{-1/2}T^{-3/2}\sum_{i=1}^N X_{i(1)}'F^0\left(F^{0\prime}F^0\right)^{-1}F^{0\prime}\left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right)F^0H_B\lambda_{i(1)}^* \\ &\quad +N^{-1/2}T^{-3/2}\sum_{i=1}^N X_{i(1)}'M_{F^0}\left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right)\left(\tilde{F}_{(1)} - F^0H_B\right)\lambda_{i(1)}^* \\ &\equiv I + II + III, \text{ say,}\end{aligned}$$

where recall $H_B = F^0H_{(1)}V_{NT,11}^{-1}$. Let $c_{K_0} \equiv (c_{1K_0}, ..., c_{K_0K_0})'$ be an arbitrary $K_0 \times 1$ nonrandom vector with $\|c_{K_0}\| = 1$. By Lemma D.4(iii) and Assumptions A.1(iii) and A.3(i),

$$\begin{aligned}\left|c_{K_0}'I\right| &= N^{-1/2}T^{-3/2}\left|\sum_{k=1}^{K_0}c_{kK_0}\operatorname{tr}\left(\mathbf{X}_k\left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right)F^0\lambda^{0\prime}\right)\right| \\ &\leq \left\{T^{-3}\sum_{k=1}^{K_0}\left\|\mathbf{X}_k\left(N^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \Phi_{NT}\right)F^0\right\|^2\right\}^{1/2}\left\{N^{-1/2}\left\|\lambda^0\right\|\right\} \\ &= K_0^{1/2}\left[O_P\left(N^{-1/2}\right) + o_P\left(T^{-(7+4\sigma)/(16+8\sigma)}\right)\right]O_P\left(1\right) = o_P\left(1\right).\end{aligned}$$

It follows that $\|I\| = o_P(1)$. Similarly, by Lemmas D.4(i) and A.2(ii) and using $\nu_{1NT} = O_P(NT^{1/2})$, we can show that $\|II\| = O_P(N^{1/2}T^{-1} + T^{-1/2}) = o_P(1)$ and $\|III\| = \delta_{NT}^{-1}O_P(1 + N^{1/2}T^{-1/2}) = o_P(1)$.

Consequently, $\mathbb{B}_{2NT}-\bar{\mathbb{B}}_{2NT}=\mathbf{o}_P(1)$ and (ii1) follows. To show (ii2), note that $\hat{\mathbb{B}}_{2NT}-\bar{\mathbb{B}}_{2NT}=N^{-1/2}T^{-3/2}\sum_{i=1}^{N}X_{i(1)}'(M_{\hat{F}_{(1)}}\hat{\Phi}_{NT}\tilde{F}_{(1)}\hat{\lambda}_{i(1)}-M_{F^0}\Phi_{NT}\tilde{F}_{(1)}\lambda_{i(1)}^*)=DB_{2NT,1}+DB_{2NT,2}+DB_{2NT,3}$, where

$$
\begin{aligned}
DB_{2NT,1} &= N^{-1/2}T^{-3/2}\sum_{i=1}^{N}X_{i(1)}'\left(M_{\hat{F}_{(1)}}-M_{F^0}\right)\hat{\Phi}_{NT}\tilde{F}_{(1)}\hat{\lambda}_{i(1)},\\
DB_{2NT,2} &= N^{-1/2}T^{-3/2}\sum_{i=1}^{N}X_{i(1)}'M_{F^0}\left(\hat{\Phi}_{NT}-\Phi_{NT}\right)\tilde{F}_{(1)}\hat{\lambda}_{i(1)},\\
DB_{2NT,3} &= N^{-1/2}T^{-3/2}\sum_{i=1}^{N}X_{i(1)}'M_{F^0}\Phi_{NT}\tilde{F}_{(1)}\left(\hat{\lambda}_{i(1)}-\lambda_{i(1)}^*\right).
\end{aligned}
$$

By Lemmas A.2(vi), D.4(ii), and Theorem 3.3, we can readily show that $\|DB_{2NT,1}\|=O_P(N^{1/2}T^{-1/2}\delta_{NT}^{-1})$, $\|DB_{2NT,2}\|=O_P(N^{1/2}T^{-1/2}\delta_{NT}^{-1}T^{1/(8+4\sigma)})$, and $\|DB_{2NT,3}\|=O_P(N^{1/2}T^{-1})$. For example,

$$
\begin{aligned}
\left|c_{K_0}'DB_{2NT,1}\right| &\leq T^{-1/2}\left\{\frac{1}{NT}\sum_{i=1}^{N}c_{K_0}'X_{i(1)}'X_{i(1)}c_{K_0}\right\}^{1/2}\\
&\quad\times\left\{\frac{1}{T}\sum_{i=1}^{N}\hat{\lambda}_{i(1)}'\tilde{F}_{(1)}'\hat{\Phi}_{NT}'\left(M_{\hat{F}_{(1)}}-M_{F^0}\right)\left(M_{\hat{F}_{(1)}}-M_{F^0}\right)\hat{\Phi}_{NT}\tilde{F}_{(1)}\hat{\lambda}_{i(1)}\right\}^{1/2}\\
&\leq T^{-1/2}\mu_{\max}\left(\frac{1}{NT}\sum_{i=1}^{N}X_{i(1)}'X_{i(1)}\right)^{1/2}\left\|P_{\hat{F}_{(1)}}-P_{F^0}\right\|\left\|\hat{\Phi}_{NT}\right\|_{\mathrm{sp}}T^{-1/2}\left\|\tilde{F}_{(1)}\right\|\nu_{3NT}^{1/2}\\
&= T^{-1/2}O_P(1)\,O_P(\delta_{NT}^{-1})O_P(1)O_P(1)O_P(N^{1/2})=O_P(N^{1/2}T^{-1/2}\delta_{NT}^{-1}),
\end{aligned}
$$

where $\nu_{3NT}=\sum_{i=1}^{N}\left\|\lambda_{i(1)}^*\right\|^2$. It follows that $\hat{\mathbb{B}}_{2NT}-\bar{\mathbb{B}}_{2NT}=\mathbf{o}_P(1)$.

Third, we prove (iii) $\hat{\mathbb{B}}_{3NT}-\mathbb{B}_{3NT}=\mathbf{o}_P(1)$. Let $\bar{\mathbb{B}}_{3NT}=(\bar{\mathbb{B}}_{3NT,1},...,\bar{\mathbb{B}}_{3NT,K_0})'$, where $\bar{\mathbb{B}}_{3NT,k}=\frac{1}{N^{3/2}T^{1/2}}\mathrm{tr}\{[F_{(1)}^{*\prime}F_{(1)}^*]^{-1}\tilde{F}_{(1)}'F^0\lambda^{0\prime}\Psi_{NT}\mathbf{X}_kF_{(1)}^*\}$. We prove (iii) by showing that (iii1) $\mathbb{B}_{3NT}-\bar{\mathbb{B}}_{3NT}=\mathbf{o}_P(1)$ and (iii2) $\hat{\mathbb{B}}_{3NT,k}-\bar{\mathbb{B}}_{3NT,k}=\mathbf{o}_P(1)$ for $k=1,...,K_0$. For (iii1), we have by Lemma D.3(vi) and A.6(i) and Assumptions A.3(i) and A.6(i)

$$
\begin{aligned}
\left|c_{K_0}'\left(\mathbb{B}_{3NT}-\bar{\mathbb{B}}_{3NT}\right)\right| &\leq \frac{1}{N^{3/2}T^{1/2}}\left|\sum_{i=1}^{K_0}c_{kK_0}\mathrm{tr}\left[\left(F_{(1)}^{*\prime}F_{(1)}^*\right)^{-1}\tilde{F}_{(1)}'F^0\lambda^{0\prime}\left(T^{-1}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'-\Psi_{NT}\right)\mathbf{X}_kF_{(1)}^*\right]\right|\\
&\leq \frac{1}{N^{3/2}T^{1/2}}\left\|\left(F_{(1)}^{*\prime}F_{(1)}^*\right)^{-1}\tilde{F}_{(1)}'F^0\right\|\left\{\sum_{i=1}^{K_0}\left\|\lambda^{0\prime}\left(T^{-1}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'-\Psi_{NT}\right)\mathbf{X}_k\right\|^2\right\}^{1/2}\left\|F_{(1)}^*\right\|\\
&= N^{-3/2}T^{-1/2}O_P(1)\,O_P(K_0^{1/2}N^{3/2}(N^{-1/2}+T^{-1/2}))O_P\left(T^{1/2}\right)\\
&= O_P(K_0^{1/2}(N^{-1/2}+T^{-1/2}))=o_P(1).
\end{aligned}
$$

For (iii2), we decompose $c'_{K_0}(\hat{\mathbb{B}}_{3NT}-\bar{\mathbb{B}}_{3NT})$ as follows

$$\begin{aligned}
c'_{K_0}(\hat{\mathbb{B}}_{3NT}-\bar{\mathbb{B}}_{3NT}) &= \frac{1}{N^{3/2}T^{1/2}}\sum_{i=1}^{K_0} c_{kK_0}\mathrm{tr}\left\{\left[\left(\hat{F}'_{(1)}\hat{F}_{(1)}\right)^{-1}\tilde{F}'_{(1)}\hat{F}-\left(F^{*\prime}_{(1)}F^{*}_{(1)}\right)^{-1}\tilde{F}'_{(1)}F^0\right]\lambda^{0\prime}\Psi_{NT}\mathbf{X}_kF^*_{(1)}\right\} \\
&\quad +\frac{1}{N^{3/2}T^{1/2}}\sum_{i=1}^{K_0} c_{kK_0}\mathrm{tr}\left\{\left(\hat{F}'_{(1)}\hat{F}_{(1)}\right)^{-1}\tilde{F}'_{(1)}\hat{F}\left[\hat{\lambda}'\hat{\Psi}_{NT}\mathbf{X}_k\hat{F}_{(1)}-H^{+}\lambda^{0\prime}\Psi_{NT}\mathbf{X}_kF^*_{(1)}\right]\right\} \\
&\equiv DB_{3NT,1}+DB_{3NT,2}.
\end{aligned}$$

By Lemmas A.2(i) and (v), and D.3(v), $|DB_{3NT,1}|\le O_P\left(\delta_{NT}^{-2}\right)R_0\max_{1\le i\le N}\psi_{iT}N^{-3/2}T^{-1/2}\left\|H^{+}\lambda^{0\prime}\right\|$ $\times||F^*_{(1)}||\{\sum_{i=1}^{K_0}\|\mathbf{X}_k\|^2\}^{1/2}=O_P(K_0^{1/2}N^{-1/2}T^{1/2}\delta_{NT}^{-2})$, and similarly $|DB_{3NT,2}|=O_P(K_0^{1/2}N^{-1/2}T^{1/2}\delta_{NT}^{-1})$. It follows that $\hat{\mathbb{B}}_{3NT}-\bar{\mathbb{B}}_{3NT}=\mathbf{o}_P(1)$ and $\hat{\mathbb{B}}_{3NT}-\mathbb{B}_{3NT}=\mathbf{o}_P(1)$.

Fourth, we prove (iv) $\hat{\mathbb{B}}_{4NT}-\mathbb{B}_{4NT}=\mathbf{o}_P(1)$. Let $\bar{\mathbb{B}}_{4NT}=(\bar{\mathbb{B}}_{4NT,1},...,\bar{\mathbb{B}}_{4NT,K_0})'$, where $\bar{\mathbb{B}}_{4NT,k}\equiv E_{\mathcal{D}}(\mathbb{B}_{4NT,k})=\frac{1}{\sqrt{NT}}\mathrm{tr}\left[P_{F^0}E_{\mathcal{D}}\left(\boldsymbol{\varepsilon}'\bar{\mathbf{X}}_k\right)\right]$. We prove (iv) by showing that (iv1) $\mathbb{R}_{4NT}\equiv\mathbb{B}_{4NT}-\bar{\mathbb{B}}_{4NT}=\mathbf{o}_P(1)$, and (iv2) $\hat{\mathbb{B}}_{4NT}-\bar{\mathbb{B}}_{4NT}=\mathbf{o}_P(1)$. We first show (iv1). Note that the kth element of $\mathbb{R}_{4NT}$ is given by $\mathbb{R}_{4NT,k}=(NT)^{-1/2}\mathrm{tr}\{P_{F^0}\left[\boldsymbol{\varepsilon}'\mathbf{X}_k-E_{\mathcal{D}}(\boldsymbol{\varepsilon}'\mathbf{X}_k)-\boldsymbol{\varepsilon}'E_{\mathcal{D}}(\mathbf{X}_k)\right]\}$. Apparently, $E_{\mathcal{D}}(\mathbb{R}_{4NT,k})=0$. Let ι_{kK_0} be a $K_0\times1$ unit vector with 1 in its kth position and zeros elsewhere. By Assumption B.2(ii) and Jensen inequality,

$$\begin{aligned}
E_{\mathcal{D}}\left(\mathbb{R}^2_{4NT,k}\right) &= \mathrm{Var}_{\mathcal{D}}(\mathbb{R}_{4NT,k})=(NT)^{-1}E_{\mathcal{D}}\left(\sum_{i=1}^{N}\iota'_{kK_0}\left\{\left[X'_{i(1)}-E_{\mathcal{D}}\left(X'_{i(1)}\right)\right]P_{F^0}\varepsilon_i-E_{\mathcal{D}}\left(X'_{i(1)}P_{F^0}\varepsilon_i\right)\right\}\right)^2 \\
&= (NT)^{-1}\sum_{i=1}^{N}E_{\mathcal{D}}\left[\iota'_{kK_0}\left\{\left[X'_{i(1)}-E_{\mathcal{D}}\left(X'_{i(1)}\right)\right]P_{F^0}\varepsilon_i-E_{\mathcal{D}}\left(X'_{i(1)}P_{F^0}\varepsilon_i\right)\right\}\right]^2 \\
&\le (NT)^{-1}\sum_{i=1}^{N}E_{\mathcal{D}}\left[\iota'_{kK_0}\left[X'_{i(1)}-E_{\mathcal{D}}\left(X'_{i(1)}\right)\right]P_{F^0}\varepsilon_i\right]^2\equiv\Xi_{kNT},\ \text{say}.
\end{aligned}$$

Let $\eta_{ts}\equiv F_t^{0\prime}\left(T^{-1}F^{0\prime}F^0\right)^{-1}F_s^0$ and $\bar{X}_{it,k}\equiv X_{it,k}-E_{\mathcal{D}}(X_{it,k})$ for $k=1,...,K_0$. Then we can write

$$\Xi_{kNT}=\frac{1}{NT^3}\sum_{i=1}^{N}E_{\mathcal{D}}\left[\sum_{t=1}^{T}\sum_{s=1}^{T}\eta_{ts}\bar{X}_{it,k}\varepsilon_{is}\right]^2=\frac{1}{NT^3}\sum_{i=1}^{N}\sum_{t_1=1}^{T}\sum_{t_1=1}^{T}\sum_{t_3=1}^{T}\sum_{t_4=1}^{T}\eta_{t_1t_2}\eta_{t_3t_4}E_{\mathcal{D}}\left[\bar{X}_{it_1,k}\bar{X}_{it_3,k}\varepsilon_{it_2}\varepsilon_{it_4}\right].$$

Let $S\equiv\{t_1,t_2,t_3,t_4\}$. We consider three cases for the time indices in S: (a) $\#S=4$, (b) $\#S=3$, (c) $\#S\le2$. We use $E\Xi_{k,a}$, $E\Xi_{k,b}$, and $E\Xi_{k,c}$ to denote Ξ_{kNT} when the time indices in the above summation are restricted to cases (a), (b), and (c), respectively. It is easy to show that $E\Xi_{k,c}=O_P\left(T^{-1}\right)=o_P\left(K_0^{-1}\right)$. In case (a), we consider two subcases: $(a1)$ for at least two $j\in\{1,2,3,4\}$, $|t_j-t_k|\ge\tau_*$ for any $k\ne j$ and $k\in\{1,2,3,4\}$, and $(a2)$ all the other remaining cases. We use $E\Xi_{k,a1}$ and $E\Xi_{k,a2}$ to denote $E\Xi_{k,a}$ when the time indices in its summation are restricted to subcases $(a1)$ and $(a2)$, respectively. In subcase $(a1)$, wlog we assume that $t_4<t_3<t_2<t_1$. (Note if either t_4 or t_2 is largest in S, then $E_{\mathcal{D}}\left[\bar{X}_{it_1,k}\bar{X}_{it_2,k}\varepsilon_{it_2}\varepsilon_{it_4}\right]=0$ by Assumption B.2(iii).) It is easy to see that either t_4 or t_1 (or both) has to lie at least τ_*-apart from other time indices in S. Wlog, assume that t_4 lies at least τ-apart from

(t_3, t_2, t_1). Then by Davydov inequality

$$\left|E_{\mathcal{D}}\left[\bar{X}_{it_1,k}\bar{X}_{it_2,k}\varepsilon_{it_2}\varepsilon_{it_4}\right]\right| \leq 8\left\|\varepsilon_{it_4}\right\|_{8+4\sigma,\mathcal{D}}\left\|\bar{X}_{it_1,k}\bar{X}_{it_2,k}\varepsilon_{it_2}\right\|_{(8+4\sigma)/3,\mathcal{D}}\alpha_{\mathcal{D}}\left(\tau_*\right)^{(1+\sigma)/(2+\sigma)}.$$

With this, one can readily show that $|E\Xi_{k,a1}| \leq O_P(T\alpha_{\mathcal{D}}(\tau_*)^{(1+\sigma)/(2+\sigma)}) = o_P\left(K_0^{-1}\right)$. In subcase $(a2)$, noting that the total number of terms in the summation of $E\Xi_{k,a2}$ is of order $O\left(NT^2\tau_*^2\right)$, we can readily show that $|E\Xi_{k,a2}| = O_P\left(T^{-1}\tau_*^2\right) = o_P\left(K_0^{-1}\right)$. Consequently, $E\Xi_{k,a} = o_P\left(K_0^{-1}\right)$. Analogously, we can show that $E\Xi_{k,b} = o_P\left(K_0^{-1}\right)$. Consequently, we have $E_{\mathcal{D}}\left|c_{K_0}'\mathbb{R}_{4NT}\right|^2 \leq \sum_{k=1}^{K_0} E_{\mathcal{D}}\left(\mathbb{R}_{4NT,k}^2\right) = o_P(1)$ and $\|\mathbb{R}_{4NT}\| = o_P(1)$ by Chebyshev inequality. Then (iv1) follows.

Now we show (iv2). Let $\hat{\mathbb{B}}_{4NT,k} - \bar{\mathbb{B}}_{4NT,k}$ denote the kth element of $\hat{\mathbb{B}}_{4NT} - \bar{\mathbb{B}}_{4NT}$. Then

$$\begin{aligned}
& \hat{\mathbb{B}}_{4NT,k} - \bar{\mathbb{B}}_{4NT,k} \\
= \ & (NT)^{-1/2}\operatorname{tr}\left[P_{\hat{F}_{(1)}}\left(\hat{\boldsymbol{\varepsilon}}'\mathbf{X}_k\right)^{\text{trunc}} - P_{F^0}E_{\mathcal{D}}\left(\boldsymbol{\varepsilon}'\mathbf{X}_k\right)\right] \\
= \ & (NT)^{-1/2}\operatorname{tr}\left[\left(P_{\hat{F}_{(1)}} - P_{F^0}\right)\left(\hat{\boldsymbol{\varepsilon}}'\mathbf{X}_k\right)^{\text{trunc}}\right] + (NT)^{-1/2}\operatorname{tr}\left\{P_{F^0}\left[\left(\hat{\boldsymbol{\varepsilon}}'\mathbf{X}_k\right)^{\text{trunc}} - E_{\mathcal{D}}\left(\boldsymbol{\varepsilon}'\mathbf{X}_k\right)\right]\right\} \\
= \ & (NT)^{-1/2}\operatorname{tr}\left[\left(P_{\hat{F}_{(1)}} - P_{F^0}\right)\left(\hat{\boldsymbol{\varepsilon}}'\mathbf{X}_k\right)^{\text{trunc}}\right] + (NT)^{-1/2}\operatorname{tr}\left\{P_{F^0}\left[E_{\mathcal{D}}\left(\boldsymbol{\varepsilon}'\mathbf{X}_k\right)^{\text{trunc}} - E_{\mathcal{D}}\left(\boldsymbol{\varepsilon}'\mathbf{X}_k\right)\right]\right\} \\
& + (NT)^{-1/2}\operatorname{tr}\left\{P_{F^0}\left[\boldsymbol{\varepsilon}'\mathbf{X}_k - E_{\mathcal{D}}\left(\boldsymbol{\varepsilon}'\mathbf{X}_k\right)\right]^{\text{trunc}}\right\} + (NT)^{-1/2}\operatorname{tr}\left[P_{F^0}\left(\hat{\boldsymbol{\varepsilon}}'\mathbf{X}_k - \boldsymbol{\varepsilon}'\mathbf{X}_k\right)^{\text{trunc}}\right] \\
\equiv \ & DB_{41,k} + DB_{42,k} + DB_{43,k} + DB_{44,k}, \text{ say.}
\end{aligned}$$

Let $DB_{4l} = (DB_{4l,1}, ..., DB_{4l,K_0})'$ for $l = 1, 2, 3, 4$. Recall that $\varepsilon_{\cdot t} \equiv (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$ and $\hat{\varepsilon}_{\cdot t} \equiv (\hat{\varepsilon}_{1t}, ..., \hat{\varepsilon}_{Nt})'$. Let $X_{\cdot s,k}$ denote the sth columns of the $N \times T$ matrices $\mathbf{X}_k$. Then $\max_s \|X_{\cdot s,k}\| = O_P\left(N^{1/2}\right)$ by Lemma D.5(iv). Then by Lemmas D.5(ii), (iv), and (ix), we have

$$\begin{aligned}
\left\|\left(\hat{\boldsymbol{\varepsilon}}'\mathbf{X}_k\right)^{\text{trunc}}\right\|_{\text{sp}} & \leq M\max_{t,s}\left|\hat{\varepsilon}_{\cdot t}'X_{\cdot s,k}\right| \leq M\max_t\|\hat{\varepsilon}_{\cdot t}\|\max_s\|X_{\cdot s,k}\| \\
& \leq M\left\{\max_t\|\varepsilon_{\cdot t}\| + \max_t\|\hat{\varepsilon}_{\cdot t} - \varepsilon_{\cdot t}\|\right\}\max_s\|X_{\cdot s,k}\| \\
& = M\left\{O_P\left(N^{1/2}\right) + o_P\left(N^{1/2}\right)\right\}O_P\left(N^{1/2}\right) = O_P(NM) \text{ uniformly in } k.
\end{aligned}$$

It follows that by Lemma A.2(vi) and Assumption B.2(iv)

$$\begin{aligned}
\left|c_{K_0}'DB_{41}\right| & \leq R_0(NT)^{-1/2}\left\|P_{\hat{F}_{(1)}} - P_{F^0}\right\|_{\text{sp}}\left\{\sum_{k=1}^{K_0}\left\|\left(\hat{\boldsymbol{\varepsilon}}'\mathbf{X}_k\right)^{\text{trunc}}\right\|_{\text{sp}}^2\right\}^{1/2} \\
& = (NT)^{-1/2}O_P\left(\delta_{NT}^{-1}\right)O_P(K_0^{1/2}NM) = O_P\left(K_0^{1/2}MN^{1/2}T^{-1/2}\delta_{NT}^{-1}\right) = o_P(1).
\end{aligned}$$

Following Moon and Weidner (2014b), let $A = \boldsymbol{\varepsilon}'\mathbf{X}_k$ and $B = A - A^{\text{trunc}}$. Let A_{ts} and B_{ts} denote the (t, s)th elements of A and B, respectively. Then $B_{ts} = 0$ for $t < s \leq t + M$ and $B_{ts} = A_{ts}$ otherwise. By construction, $A_{ts} = 0$ for $t \geq s$. When $t < s$, we apply Davydov inequality for conditional strong mixing processes to obtain

$$|A_{ts}| = \left|\sum_{i=1}^N E_{\mathcal{D}}\left(\varepsilon_{it}X_{is,k}\right)\right| \leq \sum_{i=1}^N\left|E_{\mathcal{D}}\left(\varepsilon_{it}X_{is,k}\right)\right| \leq \sum_{i=1}^N\|\varepsilon_{it}\|_{8+4\sigma,\mathcal{D}}\|X_{is,k}\|_{8+4\sigma,\mathcal{D}}\alpha_{\mathcal{D}}(s-t)^{(3+2\sigma)/(4+2\sigma)}.$$

For an $m\times n$ matrix $E=(E_{ij})$, define $\|E\|_1 \equiv \max_{1\le j\le n}\sum_{i=1}^m |E_{ij}|$ and $\|E\|_\infty \equiv \max_{1\le i\le m}\sum_{j=1}^n |E_{ij}|$. Then

$$\begin{aligned}
\left\|E_{\mathcal{D}}\left(\boldsymbol{\varepsilon}'\mathbf{X}_k\right)^{\text{trunc}}-E_{\mathcal{D}}\left(\boldsymbol{\varepsilon}'\mathbf{X}_k\right)\right\|_1 &= \max_{1\le t\le T-M-1}\sum_{s=t+M+1}^{T}|A_{ts}| \\
&\le \max_{1\le t\le T-M-1}\sum_{i=1}^{N}\sum_{s=t+M+1}^{T}\|\varepsilon_{it}\|_{8+4\sigma,\mathcal{D}}\|X_{is,k}\|_{8+4\sigma,\mathcal{D}}\,\alpha_{\mathcal{D}}(s-t)^{(3+2\sigma)/(4+2\sigma)} \\
&\le \max_t \frac{T}{2}\sum_{i=1}^{N}\left(\|\varepsilon_{it}\|_{8+4\sigma,\mathcal{D}}^2+\|X_{it,k}\|_{8+4\sigma,\mathcal{D}}^2\right)\alpha_{\mathcal{D}}(M+1)^{(3+2\sigma)/(4+2\sigma)} \\
&= O_P\left(NT\alpha_{\mathcal{D}}(M+1)^{(3+2\sigma)/(4+2\sigma)}\right) \text{ uniformly in } k.
\end{aligned}$$

Similarly, we can show that $\left\|E_{\mathcal{D}}\left(\boldsymbol{\varepsilon}'\mathbf{X}_k\right)^{\text{trunc}}-E_{\mathcal{D}}\left(\boldsymbol{\varepsilon}'\mathbf{X}_k\right)\right\|_\infty = O_P(NT\alpha_{\mathcal{D}}(M+1)^{(3+2\sigma)/(4+2\sigma)})$ uniformly in k. By the inequality $\|E\|_{\text{sp}}^2\le\|E\|_1\|E\|_\infty$ and Assumption B.2(iv),

$$\begin{aligned}
\left|c_{K_0}'DB_{42}\right|^2 &\le \sum_{k=1}^{K_0}\left\{R_0(NT)^{-1/2}\left\|E_{\mathcal{D}}\left(\boldsymbol{\varepsilon}'\mathbf{X}_k\right)^{\text{trunc}}-E_{\mathcal{D}}\left(\boldsymbol{\varepsilon}'\mathbf{X}_k\right)\right\|_{\text{sp}}\right\}^2 \\
&\le \sum_{k=1}^{K_0}R_0^2(NT)^{-1}\left\|E_{\mathcal{D}}\left(\boldsymbol{\varepsilon}'\mathbf{X}_k\right)^{\text{trunc}}-E_{\mathcal{D}}\left(\varepsilon'\mathbf{X}_k\right)\right\|_1\left\|E_{\mathcal{D}}\left(\boldsymbol{\varepsilon}'\mathbf{X}_k\right)^{\text{trunc}}-E_{\mathcal{D}}\left(\boldsymbol{\varepsilon}'\mathbf{X}_k\right)\right\|_\infty \\
&\le O_P\left(K_0NT\alpha_{\mathcal{D}}(M+1)^{(3+2\sigma)/(2+\sigma)}\right)=o_P(1).
\end{aligned}$$

In addition, by Lemmas D.5(iv)-(v) and (viii) and Assumption B.2(iv)

$$\begin{aligned}
\left|c_{K_0}'DB_{43}\right| &\le R_0(NT)^{-1/2}\left\{\sum_{k=1}^{K_0}\left\|\left(\boldsymbol{\varepsilon}'\mathbf{X}_k\right)^{\text{trunc}}-E_{\mathcal{D}}\left(\boldsymbol{\varepsilon}'\mathbf{X}_k\right)^{\text{trunc}}\right\|_{\text{sp}}^2\right\}^{1/2} \\
&\le R_0(NT)^{-1/2}K_0^{1/2}M\max_t\max_{t<s\le t+M}\left|\varepsilon_{\cdot t}'X_{\cdot s,k}-E_{\mathcal{D}}\left(\varepsilon_{\cdot t}'X_{\cdot s,k}\right)\right| \\
&= R_0(NT)^{-1/2}K_0^{1/2}MO_P(N\alpha_{NT})=O_P\left(K_0^{1/2}N^{1/2}T^{-1/2}M\alpha_{NT}\right)=o_P(1),
\end{aligned}$$

and

$$\begin{aligned}
\left|c_{K_0}'DB_{44}\right| &\le R_0(NT)^{-1/2}\left\{\sum_{k=1}^{K_0}\left\|\left(\hat{\boldsymbol{\varepsilon}}'\mathbf{X}_k\right)^{\text{trunc}}-\left(\boldsymbol{\varepsilon}'\mathbf{X}_k\right)^{\text{trunc}}\right\|_{\text{sp}}^2\right\}^{1/2} \\
&\le R_0(NT)^{-1/2}K_0^{1/2}M\max_t\max_{t<s\le t+M}\left|\left(\hat{\varepsilon}_{\cdot t}-\varepsilon_{\cdot t}\right)'X_{\cdot s,k}\right| \\
&\le R_0(NT)^{-1/2}K_0^{1/2}M\max_{i,t}|\hat{\varepsilon}_{it}-\varepsilon_{it}|\max_s\|X_{\cdot s,k}\| \\
&= R_0(NT)^{-1/2}K_0^{1/2}Mo_P(1)\max_t O_P\left(N^{1/2}\right)=o_P\left(K_0^{1/2}T^{-1/2}M\right)=o_P(1).
\end{aligned}$$

Consequently, $\hat{\mathbb{B}}_{4NT}-\bar{\mathbb{B}}_{4NT}=\mathbf{o}_P(1)$ and (iv) follows.

Finally, we prove (v) $\left\|\hat{\mathbb{V}}_{NT}-\mathbb{V}_{NT}\right\|_{\text{sp}}=o_P(1)$. Noting that

$$\begin{aligned}
\hat{\mathbb{V}}_{NT}-\mathbb{V}_{NT} &= \hat{D}_{\hat{F}_{(1)}}^{-1}\hat{\Theta}_{NT}\hat{D}_{\hat{F}_{(1)}}^{-1}-D_{F^0}^{-1}\Theta_{NT}D_{F^0}^{-1} \\
&= \left(\hat{D}_{\hat{F}_{(1)}}^{-1}-D_{F^0}^{-1}\right)\hat{\Theta}_{NT}\hat{D}_{\hat{F}_{(1)}}^{-1}+D_{F^0}^{-1}\left(\hat{\Theta}_{NT}-\Theta_{NT}\right)D_{F^0}^{-1}+D_{F^0}^{-1}\Theta_{NT}\left(\hat{D}_{\hat{F}_{(1)}}^{-1}-D_{F^0}^{-1}\right),
\end{aligned}$$

the conclusion follows provided (v1) $\left\|\hat{\Theta}_{NT}-\Theta_{NT}\right\|_{\mathrm{sp}}=o_P(1)$, (v2) $\left\|\hat{D}_{\hat{F}_{(1)}}-D_{F^0}\right\|_{\mathrm{sp}}=o_P(1)$, and (v3) the eigenvalues of D_{F^0} and Θ_{NT} are uniformly bounded away from zero and infinity as $(N,T)\to\infty$. As stated in the proof of Theorem 3.3, it is trivial to show that $\left\|\hat{D}_{\hat{F}_{(1)}}-D_{F^0}\right\|_{\mathrm{sp}}\leq\left\|\hat{D}_{\hat{F}_{(1)}}-D_{F^0}\right\|=o_P(1)$. (v3) is ensured by Assumptions A.4(i) and A.5(i). We are left to show (v1). We decompose $\hat{\Theta}_{NT}-\Theta_{NT}$ as follows:

$$\begin{aligned}\hat{\Theta}_{NT}-\Theta_{NT} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it}^2-\varepsilon_{it}^2\right)\bar{Z}_{it}\bar{Z}_{it}'+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}^2\left(\hat{Z}_{it}\hat{Z}_{it}'-\bar{Z}_{it}\bar{Z}_{it}'\right)\\ &\quad+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it}^2-\varepsilon_{it}^2\right)\left(\hat{Z}_{it}\hat{Z}_{it}'-\bar{Z}_{it}\bar{Z}_{it}'\right)\equiv\Theta_{NT,1}+\Theta_{NT,2}+\Theta_{NT,3}.\end{aligned}$$

It suffices to prove $\|\Theta_{NT,s}\|_{\mathrm{sp}}=o_P(1)$ for $s=1,2,3$. Let c_{K_0} be an arbitrary $K\times 1$ nonrandom vectors with $\|c_{K_0}\|=1$. By the fact that $a^2-b^2=(a-b)^2+2(a-b)b$ and the triangle inequality,

$$\begin{aligned}\left|c_{K_0}'\Theta_{NT,1}c_{K_0}\right| &= \left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it}^2-\varepsilon_{it}^2\right)c_{K_0}'\bar{Z}_{it}\bar{Z}_{it}'c_{K_0}\right|\\ &\leq \left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it}^2-\varepsilon_{it}^2\right)c_{K_0}'\bar{Z}_{it}\bar{Z}_{it}'c_{K_0}\right|+2\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it}-\varepsilon_{it}\right)\varepsilon_{it}c_{K_0}'\bar{Z}_{it}\bar{Z}_{it}'c_{K_0}\right|.\end{aligned}$$

The first term is bounded from above by $\max_{1\leq i\leq N}\max_{1\leq t\leq T}\left(\hat{\varepsilon}_{it}-\varepsilon_{it}\right)^2\mu_{\max}(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{Z}_{it}\bar{Z}_{it}')=o_P(1)\,O_P(1)=o_P(1)$ and Lemma D.5(viii) and Assumption A.5(i). For the second term, using the expansion of $\hat{\varepsilon}_{it}-\varepsilon_{it}$ in (D.4), we can readily show that $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it}-\varepsilon_{it}\right)\varepsilon_{it}c_{K_0}'\bar{Z}_{it}\bar{Z}_{it}'c_{K_0}=o_P(1)$ uniformly in c_{K_0}. It follows that $\|\Theta_{NT,1}\|_{\mathrm{sp}}=o_P(1)$.

To prove that $\|\Theta_{NT,s}\|_{\mathrm{sp}}=o_P(1)$ for $s=2,3$, we argue that it suffices to show that uniformly in c_{K_0}

$$DZ_{NT}^{(r)}\equiv\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\varepsilon_{it}|^r\left|c_{K_0}'\left(\hat{Z}_{it}\hat{Z}_{it}'-\bar{Z}_{it}\bar{Z}_{it}'\right)c_{K_0}\right|=o_P(1)\ \text{ for } r=0,1,2. \tag{E.1}$$

Note that (E.1) implies that $c_{K_0}'\Theta_{NT,2}c_{K_0}=o_P(1)$ by taking $r=2$ and hence $\Theta_{NT,2}=o_P(1)$. In addition,

$$\begin{aligned}\left|c_{K_0}'\Theta_{NT,3}c_{K_0}\right| &\leq \frac{1}{NT}\left|\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it}-\varepsilon_{it}\right)^2c_{K_0}'\left(\hat{Z}_{it}\hat{Z}_{it}'-\bar{Z}_{it}\bar{Z}_{it}'\right)c_{K_0}\right|\\ &\quad+\frac{2}{NT}\left|\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it}-\varepsilon_{it}\right)\varepsilon_{it}c_{K_0}'\left(\hat{Z}_{it}\hat{Z}_{it}'-\bar{Z}_{it}\bar{Z}_{it}'\right)c_{K_0}\right|\\ &\leq \max_{i,t}\left(\hat{\varepsilon}_{it}-\varepsilon_{it}\right)^2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left|c_{K_0}'\left(\hat{Z}_{it}\hat{Z}_{it}'-\bar{Z}_{it}\bar{Z}_{it}'\right)c_{K_0}\right|\\ &\quad+2\max_{i,t}\left|\hat{\varepsilon}_{it}-\varepsilon_{it}\right|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\varepsilon_{it}|\left|c_{K_0}'\left(\hat{Z}_{it}\hat{Z}_{it}'-\bar{Z}_{it}\bar{Z}_{it}'\right)c_{K_0}\right|\\ &= o_P(1)\,o_P(1)+o_P(1)\,o_P(1)=o_P(1)\end{aligned}$$

by Lemma D.5(viii) and by taking $r=0$ and 1 in (E.1). It follows that $\|\Theta_{NT,3}\|_{\mathrm{sp}}=o_P(1)$.

Now we show (E.1). We observe that

$$\begin{aligned}
DZ_{NT}^{(r)} &\leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\varepsilon_{it}|^{r}\left|c_{K_0}'\left(\hat{Z}_{it}-\bar{Z}_{it}\right)\left(\hat{Z}_{it}-\bar{Z}_{it}\right)'c_{K_0}\right| \\
&\quad +\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\varepsilon_{it}|^{r}\left|c_{K_0}'\left(\hat{Z}_{it}-\bar{Z}_{it}\right)\bar{Z}_{it}'c_{K_0}\right| \equiv \varpi_{1NT}^{(r)}+\varpi_{2NT}^{(r)}, \text{ say.}
\end{aligned}$$

Notice that

$$\begin{aligned}
\hat{Z}_i-\bar{Z}_i &= -\left[P_{\hat{F}_{(1)}}X_{i(1)}-P_{F^0}E_{\mathcal{D}}\left(X_{i(1)}\right)\right]-\left[M_{\hat{F}_{(1)}}\hat{\mathcal{X}}_{i1NT}-M_{F^0}\bar{\mathcal{X}}_{i1NT}\right] \\
&\quad +X_i\left(\hat{W}_{NT}^{-1}\hat{C}_{NT}-W_0^{-1}C_0\right)-\left[P_{\hat{F}_{(1)}}X_i\hat{W}_{NT}^{-1}\hat{C}_{NT}-P_{F^0}E_{\mathcal{D}}(X_i)W_0^{-1}C_0\right] \\
&\quad -\left[M_{\hat{F}_{(1)}}\hat{\mathcal{X}}_{i2NT}\hat{W}_{NT}^{-1}\hat{C}_{NT}-M_{F^0}\bar{\mathcal{X}}_{i2NT}W_0^{-1}C_0\right] \\
&\equiv -\chi_{1i}-\chi_{2i}+\chi_{3i}-\chi_{4i}-\chi_{5i}, \text{ say.}
\end{aligned}$$

Let χ_{lit}' denote the tth row of χ_{li} for $l=1,2,...,5,$ and $t=1,2,...,T.$ Note that $\varpi_{1NT}^{(r)}(c_{K_0}) \leq 5\sum_{l=1}^{5}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\varepsilon_{it}|^{r}\ c_{K_0}'\chi_{lit}\chi_{lit}'c_{K_0} \equiv 5\sum_{l=1}^{5}\chi_{l,NT}^{(r)}$, say. We show that $\varpi_{1NT}^{(r)}=o_P(1)$ by showing that $\chi_{l,NT}^{(r)}=o_P(1)$ for $l=1,2,...,5.$

Let ι_{tT} denote a $T\times 1$ unit vector with one in its tth position and zeros elsewhere. Noting that $\chi_{1it}=[X_{i(1)}'P_{\hat{F}_{(1)}}-E_{\mathcal{D}}(X_{i(1)}')P_{F^0}]\iota_{tT}=[X_{i(1)}'-E_{\mathcal{D}}(X_{i(1)}')]P_{F^0}\iota_{tT}+E_{\mathcal{D}}(X_{i(1)}')(P_{\hat{F}_{(1)}}-P_{F^0})\iota_{tT}+[X_{i(1)}'-E_{\mathcal{D}}(X_{i(1)}')](P_{\hat{F}_{(1)}}-P_{F^0})\iota_{tT}$, we have

$$\begin{aligned}
\chi_{1,NT}^{(r)} &\leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\varepsilon_{it}|^{r}\left\|\chi_{1it}\right\|^2 \\
&\leq \frac{3}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\varepsilon_{it}|^{r}\left\|\left[X_{i(1)}'-E_{\mathcal{D}}\left(X_{i(1)}'\right)\right]P_{F^0}\iota_{tT}\right\|^2 \\
&\quad +\frac{3}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\varepsilon_{it}|^{r}\left\|E_{\mathcal{D}}\left(X_{i(1)}'\right)\left(P_{\hat{F}_{(1)}}-P_{F^0}\right)\iota_{tT}\right\|^2 \\
&\quad +\frac{3}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\varepsilon_{it}|^{r}\left\|\left[X_{i(1)}'-E_{\mathcal{D}}\left(X_{i(1)}'\right)\right]\left(P_{\hat{F}_{(1)}}-P_{F^0}\right)\iota_{tT}\right\|^2 \\
&\equiv 3\chi_{1,NT1}^{(r)}+3\chi_{1,NT2}^{(r)}+3\chi_{1,NT3}^{(r)}, \text{ say.}
\end{aligned}$$

We want to show that $\chi_{1,NT}^{(r)}=o_P(1)$ for $r=0,1,2,4$ by showing that $\chi_{1,NTl}^{(r)}=o_P(1)$ for $l=1,2,3$ and $r=0,1,2,4.$ For $\chi_{1,NT1}^{(r)}$, we have by Lemma D.2(vii)

$$\begin{aligned}
\chi_{1,NT1}^{(r)} &\leq \left\|\left(T^{-1}F^{0\prime}F^0\right)^{-1}\right\|^2\frac{1}{NT}\sum_{i=1}^{N}\left\|T^{-1}\left[X_{i(1)}'-E_{\mathcal{D}}\left(X_{i(1)}'\right)\right]F^0\right\|^2\sum_{t=1}^{T}|\varepsilon_{it}|^{r}\left\|F_t^0\right\|^2 \\
&\leq \max_i\left\|T^{-1}\left[X_{i(1)}'-E_{\mathcal{D}}\left(X_{i(1)}'\right)\right]F^0\right\|^2\left\|\left(T^{-1}F^{0\prime}F^0\right)^{-1}\right\|^2\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\varepsilon_{it}|^{r}\left\|F_t^0\right\|^2\right\} \\
&= O_P\left(K_0\alpha_{NT}^2\right)O_P(1)\,O_P(1)=o_P(1)\,.
\end{aligned}$$

For $\chi^{(r)}_{1,NT2}$, we have by Lemmas A.2(i) and (v) and Lemma D.5(vi)

$$\begin{aligned}
\chi^{(r)}_{1,NT2} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\varepsilon_{it}|^r \left\| E_{\mathcal{D}}\left(X'_{i(1)}\right)\left[\hat{F}_{(1)}\left(\hat{F}'_{(1)}\hat{F}_{(1)}\right)^{-1}\hat{F}_{t(1)} - F^*_{(1)}\left(F^{*\prime}_{(1)}F^*_{(1)}\right)^{-1}F^*_{t(1)}\right]\right\|^2 \\
&\leq \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\varepsilon_{it}|^r \left\| E_{\mathcal{D}}\left(X'_{i(1)}\right)\left[\hat{F}_{(1)}\left(\hat{F}'_{(1)}\hat{F}_{(1)}\right)^{-1} - F^*_{(1)}\left(F^{*\prime}_{(1)}F^*_{(1)}\right)^{-1}\right]F^*_{t(1)}\right\|^2 \\
&\quad + \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\varepsilon_{it}|^r \left\| E_{\mathcal{D}}\left(X'_{i(1)}\right)\hat{F}_{(1)}\left(\hat{F}'_{(1)}\hat{F}_{(1)}\right)^{-1}\left(\hat{F}_{t(1)} - F^*_{t(1)}\right)\right\|^2 \\
&\leq \left\{T\left\|\hat{F}_{(1)}\left(\hat{F}'_{(1)}\hat{F}_{(1)}\right)^{-1} - F^*_{(1)}\left(F^{*\prime}_{(1)}F^*_{(1)}\right)^{-1}\right\|^2\right\}\left\{\frac{2}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}|\varepsilon_{it}|^r\left\|E_{\mathcal{D}}\left(X'_{i(1)}\right)\right\|^2\left\|F^*_{t(1)}\right\|^2\right\} \\
&\quad + \left\{T\left\|\hat{F}_{(1)}\left(\hat{F}'_{(1)}\hat{F}_{(1)}\right)^{-1}\right\|^2\right\}\max_t\left\|\hat{F}_{t(1)} - F^*_{t(1)}\right\|^2\left\{\frac{2}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}|\varepsilon_{it}|^r\left\|E_{\mathcal{D}}\left(X'_{i(1)}\right)\right\|^2\right\} \\
&= O_P\left(\delta_{NT}^{-2}\right)O_P\left(K_0\right) + O_P\left(1\right)O_P\left(\eta_{1NT}\right)O_P\left(K_0\right) = o_P\left(1\right).
\end{aligned}$$

Analogously we can show that $\chi^{(r)}_{1,NT3} = o_P(1)$. It follows that $\chi^{(r)}_{1,NT} = o_P(1)$ for $r = 0, 1, 2, 4$.

Next, noting that $\chi_{2i} = M_{F^0}\left(\hat{\mathcal{X}}_{i1NT} - \bar{\mathcal{X}}_{i1NT}\right) + (M_{\hat{F}_{(1)}} - M_{F^0})\bar{\mathcal{X}}_{i1NT} + (M_{\hat{F}_{(1)}} - M_{F^0})\left(\hat{\mathcal{X}}_{i1NT} - \bar{\mathcal{X}}_{i1NT}\right)$, we have

$$\begin{aligned}
\chi^{(r)}_{2,NT} &\leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\varepsilon_{it}|^r\left\|\chi_{2it}\right\|^2 \\
&\leq \frac{3}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\varepsilon_{it}|^r\left\|\left(\hat{\mathcal{X}}_{i1NT} - \bar{\mathcal{X}}_{i1NT}\right)' M_{F^0}\iota_{tT}\right\|^2 \\
&\quad + \frac{3}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\varepsilon_{it}|^r\left\|\bar{\mathcal{X}}'_{i1NT}\left(M_{\hat{F}_{(1)}} - M_{F^0}\right)\iota_{tT}\right\|^2 \\
&\quad + \frac{3}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\varepsilon_{it}|^r\left\|\left(\hat{\mathcal{X}}_{i1NT} - \bar{\mathcal{X}}_{i1NT}\right)'\left(M_{\hat{F}_{(1)}} - M_{F^0}\right)\iota_{tT}\right\|^2 \\
&\equiv 3\chi^{(r)}_{2,NT1} + 3\chi^{(r)}_{2,NT2} + 3\chi^{(r)}_{2,NT3}, \text{ say.}
\end{aligned}$$

Noting that

$$\begin{aligned}
&\hat{\mathcal{X}}_{i1NT} - \bar{\mathcal{X}}_{i1NT} \\
&= \frac{1}{NT}\sum_{j=1}^{N}\left[\hat{\lambda}'_{i(1)}\hat{F}'_{(1)}\hat{F}_{(1)}\hat{\lambda}_{j(1)} - \lambda_i^{0\prime}F^{0\prime}F^0\lambda_j^0\right]X_{jt(1)} + \frac{1}{NT}\sum_{j=1}^{N}\lambda_i^{0\prime}F^{0\prime}F^0\lambda_j^0\left[X_{jt(1)} - E_{\mathcal{D}}\left(X_{jt(1)}\right)\right] \\
&= \left(\hat{\lambda}_{i(1)} - \lambda^*_{i(1)}\right)'\left(T^{-1}F^{0\prime}F^0\right)\frac{1}{N}\sum_{j=1}^{N}\lambda_j^0 X_{jt(1)} + \hat{\lambda}'_{i(1)}T^{-1}\left(\hat{F}'_{(1)}\hat{F}_{(1)} - H'_{(1)}F^{0\prime}F^0H_{(1)}\right)\frac{1}{N}\sum_{j=1}^{N}\lambda^*_{j(1)}X_{jt(1)} \\
&\quad + \hat{\lambda}'_{i(1)}T^{-1}\hat{F}'_{(1)}\hat{F}_{(1)}\frac{1}{N}\sum_{j=1}^{N}\left(\hat{\lambda}_{j(1)} - \lambda^*_{j(1)}\right)X_{jt(1)} + \frac{1}{NT}\sum_{j=1}^{N}\lambda_i^{0\prime}F^{0\prime}F^0\lambda_j^0\left[X_{jt(1)} - E_{\mathcal{D}}\left(X_{jt(1)}\right)\right],
\end{aligned}$$

we have

$$
\begin{aligned}
\chi_{2,NT1} \leq\; & 4\left\|T^{-1}F^{0\prime}F^{0}\right\|^{2}\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\lambda}_{i(1)}-\lambda_{i(1)}^{*}\right\|\frac{1}{T}\sum_{t=1}^{T}\left|\varepsilon_{it}\right|^{r}\left\|\frac{1}{N}\sum_{j=1}^{N}\lambda_{j}^{0}X_{jt(1)}\right\|^{2} \\
& +4\left\|T^{-1}\left(\hat{F}_{(1)}^{\prime}\hat{F}_{(1)}-H_{(1)}^{\prime}F^{0\prime}F^{0}H_{(1)}\right)\right\|^{2}\frac{3}{NT}\sum_{i=1}^{N}\left\|\hat{\lambda}_{i(1)}\right\|^{2}\sum_{t=1}^{T}\left|\varepsilon_{it}\right|^{r}\left\|\frac{1}{N}\sum_{j=1}^{N}\lambda_{j(1)}^{*}X_{jt(1)}\right\|^{2} \\
& +4\left\|T^{-1}\hat{F}_{(1)}^{\prime}\hat{F}_{(1)}\right\|^{2}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left|\varepsilon_{it}\right|^{r}\left\|\hat{\lambda}_{i(1)}^{\prime}\right\|^{2}\left\|\frac{1}{N}\sum_{j=1}^{N}\left(\hat{\lambda}_{j(1)}-\lambda_{j(1)}^{*}\right)X_{jt(1)}\right\|^{2} \\
& +\frac{4}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left|\varepsilon_{it}\right|^{r}\left\|\frac{1}{NT}\sum_{j=1}^{N}\lambda_{i}^{0\prime}F^{0\prime}F^{0}\lambda_{j}^{0}\left[X_{jt(1)}-E_{\mathcal{D}}\left(X_{jt(1)}\right)\right]\right\|^{2}.
\end{aligned}
$$

One can readily show that each term on the right hand side of the last expression is $o_P(1)$. For example, the first term is $o_P(1)$ because it is bounded above by

$$
\begin{aligned}
& 4\left\|T^{-1}F^{0\prime}F^{0}\right\|^{2}\left\{\frac{1}{N}||\hat{\lambda}_{(1)}-\lambda_{(1)}^{*}||^{2}\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T}\left|\varepsilon_{it}\right|^{r}\left\|\frac{1}{N}\sum_{j=1}^{N}\lambda_{j}^{0}X_{jt(1)}\right\|^{2}\right]^{2}\right\}^{1/2} \\
= \;& O_P(1)\,O_P\left(T^{-1/2}\right)O_P(K_0)=o_P(1).
\end{aligned}
$$

To see why the last term is $o_P(1)$, we first notice that $\max_{1\leq t\leq T}\left\|\frac{1}{N}\sum_{j=1}^{N}\lambda_{j}^{0}\left[X_{jt(1)}-E_{\mathcal{D}}\left(X_{jt(1)}\right)\right]\right\|=O_P(K_0^{1/2}\alpha_{TN})$ by arguments as used in proving Lemma D.5(iii). Then

$$
\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left|\varepsilon_{it}\right|^{r}\left\|\frac{1}{NT}\sum_{j=1}^{N}\lambda_{i}^{0\prime}F^{0\prime}F^{0}\lambda_{j}^{0}\left[X_{jt(1)}-E_{\mathcal{D}}\left(X_{jt(1)}\right)\right]\right\|^{2} \\
\leq\; & O_P\left(K_0\alpha_{TN}^{2}\right)\left\|(T^{-1}F^{0\prime}F^{0})\right\|^{2}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left|\varepsilon_{it}\right|^{r}\left\|\lambda_{i}^{0}\right\|^{2}=O_P\left(K_0\alpha_{NT}^{2}\right)=o_P(1).
\end{aligned}
$$

Consequently, $\chi_{2,NT1}^{(r)}=o_P(1)$. Analogously, using Lemma A.2(vi) we can show that $\chi_{2,NTs}^{(r)}=o_P(1)$ for $s=2,3$. Hence $\chi_{2,NT}^{(r)}=o_P(1)$ for $r=0,1,2,4$.

Next, we can readily show that $\hat{W}_{NT}=W_0+\mathbf{o}_P(K^{-1/2})$ and $\hat{C}_{NT}=C_0+\mathbf{o}_P(K^{-1/2})$. It follows that for $r=0,1,2,4$

$$
\begin{aligned}
\chi_{3,NT}^{(r)} =\; & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left|\varepsilon_{it}\right|^{r}c_{K_0}^{\prime}\left(\hat{W}_{NT}^{-1}\hat{C}_{NT}-W_0^{-1}C_0\right)^{\prime}X_{it}X_{it}^{\prime}\left(\hat{W}_{NT}^{-1}\hat{C}_{NT}-W_0^{-1}C_0\right)c_{K_0} \\
\leq\; & \left\|\hat{W}_{NT}^{-1}\hat{C}_{NT}-W_0^{-1}C_0\right\|^{2}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left|\varepsilon_{it}\right|^{r}\left\|X_{it}\right\|^{2}=o_P\left(K^{-1}\right)O_P(K)=o_P(1).
\end{aligned}
$$

Analogously, we can show that $\chi_{4,NT}^{(r)}=o_P(1)$ and $\chi_{5,NT}^{(r)}=o_P(1)$ for $r=0,1,2,4$. It follows that $\bar{\varpi}_{1NT}^{(r)}(c_{K_0})=o_P(1)$ and thus $\varpi_{1NT}^{(r)}=o_P(1)$ for $r=0,1,2,4$.

By CS inequality, $\varpi_{2NT}^{(r)} \leq \{\varpi_{1NT}^{(2r)}\}^{1/2}\{\bar{\varpi}_{2NT}\}^{1/2} = o_P(1)\,O_P(1) = o_P(1)$ for $r = 0, 1, 2$, where we use the fact that $\bar{\varpi}_{2NT} \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} c_{K_0}'\bar{Z}_{it}\bar{Z}_{it}'c_{K_0} \leq \mu_{\max}(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{Z}_{it}\bar{Z}_{it}') = O_P(1)$. Consequently we have $DZ_{NT}^{(r)} = o_P(1)$ for $r = 0, 1, 2$. This completes the proof of the corollary. ■

F Justification of Assumption A.1(i)

In this appendix we justify Assumption A.1(i) by arguing that the bias-corrected initial estimator $\tilde{\beta}^c$ can be obtained as in Moon Weidner (2014b, MWb hereafter).

Let $\tilde{\beta} = \tilde{\beta}(R)$ be as defined in Section 2 when $R \geq R_0$ factors are assumed in the estimation. Let $\tilde{\beta}^c = \tilde{\beta}^c(R)$ be its bias-corrected version. Moon and Weidner (2014a, MWa hereafter) show that $\sqrt{NT}\left(\tilde{\beta}(R) - \beta^0\right)$ is asymptotically equivalent to $\sqrt{NT}\left(\tilde{\beta}(R_0) - \beta^0\right)$ where K is fixed, N and T pass to infinity at the same rate ($N/T \to \kappa^2 \in (0, \infty)$), and $R > R_0$; MWb show that $\sqrt{NT}\left(\tilde{\beta}^c(R) - \beta^0\right)$ follows asymptotic normal distribution when K is fixed, N and T pass to infinity at the same rate, and $R = R_0$. Here, we allow K to pass to infinity at a controllable rate such that Assumptions A.3 (i) and A.6(i) are satisfied but restrict R to be finite. In addition, we allow that N and T to pass to infinity at different rates.

Define
$$\mathcal{L}_{NT}^{R}(\beta) \equiv \frac{1}{NT}\sum_{r=R+1}^{T}\mu_r\left[\left(\mathbf{Y} - \sum_{k=1}^{K}\beta_k\mathbf{X}_k\right)'\left(\mathbf{Y} - \sum_{k=1}^{K}\beta_k\mathbf{X}_k\right)\right].$$
Note that the (k_1, k_2)th element of the $K \times K$ matrix W_{NT} defined in Section 3.1 is given by $W_{NT,k_1k_2} = \frac{1}{NT}\mathrm{tr}(M_{\lambda^0}\mathbf{X}_{k_1}M_{F^0}\mathbf{X}_{k_1})$. Let $\Phi \equiv \lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(F^{0\prime}F^0\right)^{-1}F^{0\prime}$. Let $C_{NT}^{(1)}$ and $C_{NT}^{(2)}$ be $K \times 1$ vectors whose kth elements are respectively given by

$$\begin{aligned}
C_{NT,k}^{(1)} &\equiv \frac{1}{\sqrt{NT}}\mathrm{tr}\left(M_{\lambda^0}\mathbf{X}_k M_{F^0}\boldsymbol{\varepsilon}'\right), \\
C_{NT,k}^{(2)} &\equiv -\frac{1}{\sqrt{NT}}\mathrm{tr}\left(\mathbf{X}_k\Phi'\boldsymbol{\varepsilon}M_{F^0}\boldsymbol{\varepsilon}'M_{\lambda^0} + \mathbf{X}_k M_{F^0}\boldsymbol{\varepsilon}'M_{\lambda^0}\boldsymbol{\varepsilon}\Phi' + \mathbf{X}_k M_{F^0}\boldsymbol{\varepsilon}'\Phi\boldsymbol{\varepsilon}'M_{\lambda^0}\right) \\
&\equiv C_{NT,k}^{(2,1)} + C_{NT,k}^{(2,2)} + C_{NT,k}^{(2,3)},
\end{aligned}$$

where $C_{NT,k}^{(2,s)}$ denotes the kth element of $C_{NT}^{(2,s)}$ for $s = 1$, 2, and 3. Let $\Delta\beta = \beta - \beta^0 = (\Delta\beta_1, ..., \Delta\beta_K)'$. Define

$$\begin{aligned}
d(\beta) = \sum_{r=1}^{R-R_0}\Bigg\{&\mu_r\left[M_{F^0}\left(\boldsymbol{\varepsilon} - \sum_{k=1}^{K}\Delta\beta_k\mathbf{X}_k\right)'M_{\lambda^0}\left(\boldsymbol{\varepsilon} - \sum_{k=1}^{K}\Delta\beta_k\mathbf{X}_k\right)M_{F^0}\right] \\
&-\mu_r\left[M_{F^0}\boldsymbol{\varepsilon}'M_{\lambda^0}\boldsymbol{\varepsilon}M_{F^0}\right] - \mu_r\left[M_{F^0}\left(\sum_{k=1}^{K}\Delta\beta_k\mathbf{X}_k\right)'M_{\lambda^0}\left(\sum_{k=1}^{K}\Delta\beta_k\mathbf{X}_k\right)M_{F^0}\right]\Bigg\}.
\end{aligned}$$

Let $B(\beta)$ be as defined in MWa (p.18 in the supplementary appendix). We make some additional assumptions.

Assumption S.1 (i) $\max_{1\leq k\leq K} E\left\|F^{0\prime}\boldsymbol{\varepsilon}'\mathbf{X}_k\right\|_{\text{sp}}^2 = O\left(NT\left(N+T\right)\right).$

(ii) $\max_{1\leq k\leq K} E\left\|\lambda^0\mathbf{X}_k\boldsymbol{\varepsilon}'\right\|_{\text{sp}}^2 = O\left(NT\left(N+T\right)\right).$

Assumption S.2 For any $L>0$, we have

(i) $\sup_{\beta:\|\beta-\beta^0\|\leq L\delta_{NT}^{-1}} \frac{\max(d(\beta),0)}{\sqrt{N+T}+(N+T)\delta_{NT}^{1/2}\|\beta-\beta^0\|+NT\|\beta-\beta^0\|^2/\ln N} = O_P\left(1\right)$, and

(ii) $\sup_{\beta:\|\beta-\beta^0\|\leq L\delta_{NT}^{-3/2}} \frac{\mu_r\left[B(\beta)+B(\beta)'\right]-\mu_r\left[B(\beta^0)+B(\beta^0)'\right]}{\left(1+\sqrt{NT}\|\beta-\beta^0\|\right)^2} = O_P\left(1\right).$

Assumption S.1 can be verified easily under Assumptions B.1 and B.2. Assumption S.2(i) is a high-level condition and parallels Assumption HL1 in MWa which incorporates the case where N and T diverge to infinity at different rates. It can be verified under some primitive conditions as specified in MWa by modifying the proof of Lemma S.8 in the latter paper. For example, if for each $k=1,...,K$, we have $\mathbf{X}_k=\mathbf{X}_k\left(1\right)+\mathbf{X}_k\left(2\right)$, where $M_{\lambda^0}\mathbf{X}_k\left(1\right)M_{F^0}=0$, $E\left\|\mathbf{X}_k\left(1\right)\right\|_{\text{sp}}^2=O\left(NT\right)$, and $E\left\|\mathbf{X}_k\left(2\right)\right\|_{\text{sp}}^2=O(\delta_{NT}^3)$. [This condition essentially reduces to Assumption DX-2 in MWa when N and T pass to infinity at the same rate.] Then we can readily verify Assumption S.2(i) under Assumptions A.1(ii)-(iii) and (v). Assumption S.2(ii) parallels Assumption HL2 in MWa and can also be verified under some primitive conditions.

Let $\tilde{\Phi}\equiv\tilde{\lambda}\left(\tilde{\lambda}'\tilde{\lambda}\right)^{-1}\left(\tilde{F}'\tilde{F}\right)^{-1}\tilde{F}'$. Define $\tilde{B}_{NT}^{(l)}=(\tilde{B}_{NT,1}^{(l)},...,\tilde{B}_{NT,K}^{(l)})'$, $l=1,2,3$, where

$$\begin{aligned}
\tilde{B}_{NT,k}^{(1)} &= \frac{1}{\sqrt{NT}}\text{tr}\left[P_{\hat{F}_{(1)}}\left(\tilde{\boldsymbol{\varepsilon}}'\mathbf{X}_k\right)^{\text{trunc}}\right], \\
\tilde{B}_{NT,k}^{(2)} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\varepsilon}_{it}^2\left[M_{\tilde{\lambda}}\mathbf{X}_k\tilde{\Phi}'\right]_{ii}, \\
\tilde{B}_{NT,k}^{(3)} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\varepsilon}_{it}^2\left[M_{\tilde{F}}\mathbf{X}_k'\tilde{\Phi}\right]_{tt},
\end{aligned}$$

where $[B]_{ij}$ denotes the (i,j)th element of the matrix B, $\tilde{\boldsymbol{\varepsilon}}$ is the residual matrix based on the initial estimators $\tilde{\beta}$, $\tilde{F}$, and $\tilde{\lambda}$ of β, F and λ with (i,t)th element given by $\tilde{\varepsilon}_{it}=Y_{it}-\tilde{\beta}'X_{it}-\tilde{\lambda}_i'\tilde{F}_t$. [Note that $\tilde{B}_{NT,k}^{(l)}$ corresponds to $\sqrt{\frac{T}{N}}\hat{B}_{l,k}$ in MWb for $l=1,3$, $\tilde{B}_{NT,k}^{(2)}$ corresponds to $\sqrt{\frac{N}{T}}\hat{B}_{l,k}$ in MWb, and $\tilde{B}_{NT,k}^{(1)}$ has the same structure as $\hat{\mathbb{B}}_{4NT,k}$ in our Section 3.3.] Define the bias corrected estimator as:

$$\tilde{\beta}^c\left(R\right)=\tilde{\beta}\left(R\right)+\left(NT\right)^{-1/2}\tilde{W}_{NT}^{-1}\left(\tilde{B}_{NT}^{(1)}+\tilde{B}_{NT}^{(2)}+\hat{B}_{NT}^{(3)}\right),$$

where $\tilde{W}_{NT}=\left(NT\right)^{-1}\sum_{i=1}^{N}\tilde{X}_i'M_{\tilde{F}}\tilde{X}_i$.

Let $\gamma_{NT}=\left(NT\right)^{-1/2}W_{NT}^{-1}(C_{NT}^{(1)}+C_{NT}^{(2)})$. We argue that in the case of divergent K, (i) $\left\|\tilde{\beta}\left(R\right)-\beta^0\right\|=O_P\left(\delta_{NT}^{-1}\right)$ for any fixed (finite) $R\geq R_0$; (ii) $\tilde{\beta}\left(R_0\right)-\beta^0=\gamma_{NT}+\mathcal{R}_{NT}^{R_0}$ with $\left\|\mathcal{R}_{NT}^{R_0}\right\|=O_P\left(\left\|\gamma_{NT}\right\|\delta_{NT}^{-1}\right)$; (iii) $\left\|\tilde{\beta}\left(R\right)-\beta^0\right\|=O_P(\delta_{NT}^{-3/2})$ for any fixed $R\geq R_0$; (iv) $\tilde{\beta}\left(R\right)-\beta^0=\gamma_{NT}+\mathcal{R}_{NT}^{R}$ with $\left\|\mathcal{R}_{NT}^{R}\right\|=o_P((NT)^{-1/2})$ for any fixed $R>R_0$, and (v) $\sqrt{NT/K}\left\|\tilde{\beta}^c\left(R\right)\right\|=O_P\left(1\right)$ and $\sqrt{NT}\tilde{\beta}_k^c\left(R\right)=O_P\left(1\right)$ for each $k=1,2,...,K$.

Step 1. We show (i) $\left\|\tilde{\beta}\left(R\right)-\beta^0\right\|=O_P\left(\delta_{NT}^{-1}\right)$ for any fixed $R>R_0$. Following the proof of Theorem

4.1 in MWa, we can readily show that under our Assumptions A.1 (v)-(vi) and (viii), and A.3(i),

$$\left\|\tilde{\beta}\left(R\right)-\beta^{0}\right\| = O_P\left(\delta_{NT}^{-1}\right) \text{ for any fixed } R \geq R_0.$$

In particular, these assumptions ensure that eqn. (S.5) in MWa continues to hold in our case despite the allowance of diverging K. See also Su and Zhang (2014) in the case of sieve estimation.

Step 2. We show (ii) $\tilde{\beta}\left(R_0\right)-\beta^{0}=\gamma_{NT}+\mathcal{R}_{NT}^{R_0}$ where $\left\|\mathcal{R}_{NT}^{R_0}\right\| = O_P\left(\left\|\gamma_{NT}\right\|\delta_{NT}^{-1}\right)$. Given the result in (i) and Assumptions A.1(iv)-(v) and A.3(i), we can readily show that

$$\sum_{k=1}^{K}\left|\beta_k-\beta_k^0\right|\frac{\left\|\mathbf{X}_k\right\|_{\mathrm{sp}}}{\sqrt{NT}}+\frac{\left\|\boldsymbol{\varepsilon}\right\|_{\mathrm{sp}}}{\sqrt{NT}} = O_P\left(K\delta_{NT}^{-1}\right)+O_P\left(\delta_{NT}^{-1}\right)=o_P\left(1\right)$$

for any $\beta=\left(\beta_1,...,\beta_K\right)'$ such that $\left\|\beta-\beta^0\right\| \leq L\delta_{NT}^{-1}$ where L is a large constant. This indicates Condition (S.34) in Lemma S.1 of MWa is satisfied under our Assumptions A.1(ii)-(iii). Then we can follow the proof of Theorem 4.2 in MWa and show that

$$\mathcal{L}_{NT}^{R_0}\left(\beta\right)=\mathcal{L}_{NT}^{R_0}\left(\beta^0\right)-2\left(NT\right)^{-1/2}\left(\beta-\beta^0\right)'\left(C_{NT}^{(1)}+C_{NT}^{(2)}\right)+\left(\beta-\beta^0\right)'W_{NT}\left(\beta-\beta^0\right)+\mathcal{L}_{NT}^{R_0,rem}\left(\beta\right).$$

where the remainder term $\mathcal{L}_{NT}^{0,rem}\left(\beta\right)$ satisfies

$$\sup_{\left\{\beta:\left\|\beta-\beta^0\right\|\leq L\delta_{NT}^{-1}\right\}}\mathcal{L}_{NT}^{R_0,rem}\left(\beta\right)=O_P\left(\delta_{NT}^{-3}\right).$$

The last probability order can be obtained from MWa's eqn (S.39) with $\left\|\beta-\beta^0\right\|^j$ and $N^{-j/2}$ replaced by δ_{NT}^{-j} for $j=1,2,3,4$. Following the proof of Corollary 4.3 in MWa or Theorem 3.1 in Su and Zhang (2014), we can show that

$$\tilde{\beta}\left(R_0\right)-\beta^0=\left(NT\right)^{-1/2}W_{NT}^{-1}\left(C_{NT}^{(1)}+C_{NT}^{(2)}\right)+\mathcal{R}_{NT}^{R_0}=\gamma_{NT}+\mathcal{R}_{NT}^{R_0},$$

where the remainder term $\mathcal{R}_{NT}^{R_0}$ satisfies $\left\|\mathcal{R}_{NT}^{R_0}\right\| = O_P\left(\left\|\gamma_{NT}\right\|\delta_{NT}^{-1}\right)$ (c.f., eqn. (A.9) in Su and Zhang (2014)).

Step 3. We show (iii) $\left\|\tilde{\beta}\left(R\right)-\beta^0\right\| = O_P(\delta_{NT}^{-3/2})$. Here we want to determine the probability order of $\left\|\tilde{\beta}\left(R\right)-\beta^0\right\|$ by following the arguments as used in the proof of Theorem S.5 in MWa. Under our Assumption A.1(viii), W_{NT} has minimum eigenvalue bounded away from zero asymptotically. Next, we want to determine the probability order of $\left\|C_{NT}^{(1)}\right\|$. Note that

$$\begin{aligned}
C_{NT,k}^{(1)} &= \left(NT\right)^{-1/2}\operatorname{tr}\left(M_{\lambda^0}\mathbf{X}_kM_{F^0}\boldsymbol{\varepsilon}'\right)\\
&= \left(NT\right)^{-1/2}\operatorname{tr}\left(\mathbf{X}_k\boldsymbol{\varepsilon}'\right)-\left(NT\right)^{-1/2}\operatorname{tr}\left(\mathbf{X}_kP_{F^0}\boldsymbol{\varepsilon}'\right)-\left(NT\right)^{-1/2}\operatorname{tr}\left(P_{\lambda^0}\mathbf{X}_k\boldsymbol{\varepsilon}'\right)+\left(NT\right)^{-1/2}\operatorname{tr}\left(P_{\lambda^0}\mathbf{X}_kP_{F^0}\boldsymbol{\varepsilon}'\right)\\
&= C_{NT,k}^{(1,1)}-C_{NT,k}^{(1,2)}-C_{NT,k}^{(1,3)}+C_{NT,k}^{(1,4)}, \text{ say.}
\end{aligned}$$

Let $C_{NT}^{(1,l)}$ be a $K \times 1$ vector whose kth element is given by $C_{NT,k}^{(1,l)}$. Then $\left\| C_{NT}^{(1)} \right\|^2 \leq 4 \sum_{l=1}^{4} \left\| C_{NT}^{(1,l)} \right\|^2$. By Assumption A.1(vi) and Markov inequality,

$$\left\| C_{NT}^{(1,1)} \right\|^2 = \sum_{k=1}^{K} \left\| C_{NT,k}^{(1,1)} \right\|^2 = \frac{1}{(NT)} \sum_{k=1}^{K} \left\| \mathrm{tr} \left(\mathbf{X}_k \boldsymbol{\varepsilon}' \right) \right\|^2 = O_P(K).$$

Using $\mathrm{tr}(A) \leq \mathrm{rank}(A) \|A\|$, $\|A\|_{\mathrm{sp}} \leq \|A\| \leq \sqrt{\mathrm{rank}(A)} \|A\|_{\mathrm{sp}}$, and the condition $\max_{1 \leq k \leq K} E \left\| F^{0\prime} \boldsymbol{\varepsilon}' \mathbf{X}_k \right\|_{\mathrm{sp}}^2 = O(NT(N+T))$ in Assumption S.1(i), we have

$$\begin{aligned}
\left\| C_{NT}^{(1,2)} \right\|^2 &= \frac{1}{(NT)^2} \sum_{k=1}^{K} \left| \mathrm{tr} \left(F^0 \left(F^{0\prime} F^0 \right)^{-1} F^{0\prime} \boldsymbol{\varepsilon}' \mathbf{X}_k \right) \right|^2 \leq \frac{R_0^3}{NT} \sum_{k=1}^{K} \left\| F^0 \left(F^{0\prime} F^0 \right)^{-1\prime} F^{0\prime} \boldsymbol{\varepsilon}' \mathbf{X}_k \right\|_{\mathrm{sp}}^2 \\
&\leq \frac{R_0^3}{T} \left\| T^{-1/2} F^0 \left(T^{-1} F^{0\prime} F^0 \right)^{-1} \right\|^2 \frac{1}{NT} \sum_{k=1}^{K} \left\| F^{0\prime} \boldsymbol{\varepsilon}' \mathbf{X}_k \right\|_{\mathrm{sp}}^2 = O_P \left(K N \delta_{NT}^{-2} \right).
\end{aligned}$$

Similarly, under the condition $\max_{1 \leq k \leq K} E \left\| \lambda^0 \mathbf{X}_k \boldsymbol{\varepsilon}' \right\|_{\mathrm{sp}}^2 = O(NT(N+T))$ in Assumption S.1(ii), we can show that $\left\| C_{NT}^{(1,3)} \right\|^2 = O_P(KN\delta_{NT}^{-2})$. Under Assumptions A.1 (ii)-(iv) and (vii),

$$\begin{aligned}
\left\| C_{NT}^{(1,4)} \right\|^2 &= \frac{1}{NT} \sum_{k=1}^{K} \left\| \mathrm{tr} \left(\mathbf{X}_k F^0 \left(F^{0\prime} F^0 \right)^{-1} F^{0\prime} \boldsymbol{\varepsilon}' \lambda^0 \left(\lambda^{0\prime} \lambda^0 \right)^{-1} \lambda^{0\prime} \right) \right\|^2 \\
&\leq \frac{R_0^3}{NT} \left\| F^0 \left(F^{0\prime} F^0 \right)^{-1} \right\|^2 \left\| \left(\lambda^{0\prime} \lambda^0 \right)^{-1} \lambda^{0\prime} \right\|^2 \left\| F^{0\prime} \boldsymbol{\varepsilon}' \lambda^0 \right\|^2 \sum_{k=1}^{K} \left\| \mathbf{X}_k \right\|_{\mathrm{sp}}^2 \\
&= (NT)^{-1} O_P \left(T^{-1} \right) O_P \left(N^{-1} \right) O_P(NT) O_P(NTK) = O_P(K).
\end{aligned}$$

Consequently, we have $\left\| C_{NT}^{(1)} \right\| = O_P \left(K^{1/2} N^{1/2} \delta_{NT}^{-1} \right)$. Similarly, we can show that $\left\| C_{NT}^{(2)} \right\| = O_P(K^{1/2} N^{1/2} \delta_{NT}^{-1})$. It follows that $\|\gamma_{NT}\| = O_P \left(K^{1/2} T^{-1/2} \delta_{NT}^{-1} \right)$.

Under Assumption S.2, we can follow the proof of Theorem S.5 in MWa and show that

$$\begin{aligned}
&\left(\tilde{\beta}(R) - \beta^0 \right)' W_{NT} \left(\tilde{\beta}(R) - \beta^0 \right) - 2 (NT)^{-1/2} \left(\tilde{\beta}(R) - \beta^0 \right)' \left[C_{NT}^{(1)} + C_{NT}^{(2)} \right] \\
\leq\ & \frac{1}{NT} \sum_{r=1}^{R-R_0} \left\{ \mu_r \left[M_{F^0} \left(\sum_{k=1}^{K} \Delta\beta_k \mathbf{X}_k \right)' M_{\lambda^0} \left(\sum_{k=1}^{K} \Delta\beta_k \mathbf{X}_k \right) M_{F^0} \right] \right. \\
&\left. + \sqrt{N+T} + (N+T) \delta_{NT}^{1/2} \left\| \tilde{\beta}(R) - \beta^0 \right\| + NT \left\| \tilde{\beta}(R) - \beta^0 \right\|^2 / \ln N \right\}.
\end{aligned}$$

Assumption A.1(viii) implies that $\mu_{\min}(W_{NT}) \geq 2b$ for some $b > 0$ in large samples. This, in conjunction with the fact that $\left\| C_{NT}^{(l)} \right\| = O_P \left(K^{1/2} N^{1/2} \delta_{NT}^{-1} \right)$ for $l = 1, 2$, implies that

$$\begin{aligned}
b \left\| \tilde{\beta}(R) - \beta^0 \right\|^2 &\leq \left\| \tilde{\beta}(R) - \beta^0 \right\| O_P \left(K^{1/2} T^{-1/2} \delta_{NT}^{-1} \right) \\
&\quad + \frac{1}{NT} \left[\sqrt{N+T} + (N+T) \delta_{NT}^{1/2} \left\| \tilde{\beta}(R) - \beta^0 \right\| + NT \left\| \tilde{\beta}(R) - \beta^0 \right\|^2 / \ln N \right],
\end{aligned}$$

or equivalently (by multiplying both sides by δ_{NT}^3/b),

$$\begin{aligned}
\delta_{NT}^3 \left\|\tilde{\beta}(R)-\beta^0\right\|^2 &\leq O_P\left(\frac{\sqrt{N+T}\delta_{NT}^3}{NT}\right)+\left\|\tilde{\beta}(R)-\beta^0\right\| O_P\left(K^{1/2}T^{-1/2}\delta_{NT}^2+\frac{(N+T)\,\delta_{NT}^2}{NT}\delta_{NT}^{3/2}\right) \\
&\quad +o_P\left(\delta_{NT}^3\left\|\tilde{\beta}(R)-\beta^0\right\|^2\right) \\
&= O_P(1)+\left\|\tilde{\beta}(R)-\beta^0\right\| O_P\left(\delta_{NT}^{3/2}\right)+o_P\left(\delta_{NT}^3\left\|\tilde{\beta}(R)-\beta^0\right\|^2\right),
\end{aligned}$$

where we use the fact that $\frac{\sqrt{N+T}\delta_{NT}^3}{NT}=O(1)$, $\frac{(N+T)\delta_{NT}^2}{NT}=O(1)$, and $K^{1/2}T^{-1/2}\delta_{NT}^2/(\delta_{NT}^{3/2})=K^{1/2}T^{-1/2}\delta_{NT}^{1/2}=o(1)$ under Assumption A.6(i). It follows that $\delta_{NT}^3\left\|\tilde{\beta}(R)-\beta^0\right\|^2=O_P(1)$. That is, $\left\|\tilde{\beta}(R)-\beta^0\right\|^2=O_P(\delta_{NT}^{-3/2})$.

Step 4. We show (iv) $\tilde{\beta}(R)-\beta^0=\gamma_{NT}+\mathcal{R}_{NT}^R$ with $\left\|\mathcal{R}_{NT}^R\right\|=o_P((NT)^{-1/2})$. Given the result in Step 3, we can apply Assumption S.2(ii) and follow the proof of Corollary S.10 in MWa and show that

$$\mathcal{L}_{NT}^R\left(\tilde{\beta}(R)\right)\leq \mathcal{L}_{NT}^{R_0}\left(\beta^0+\gamma_{NT}\right)+\frac{1}{NT}o_P\left(1+\sqrt{NT}\left\|\tilde{\beta}(R)-\beta^0\right\|\right)^2=\mathcal{L}_{NT}^{R_0}\left(\beta^0+\gamma_{NT}\right)+o_P\left(\delta_{NT}^{-3}\right).$$

Then following Step 2 and the proof of Corollary 4.3 in MWa or Theorem 3.1 in Su and Zhang (2014), we can show that

$$\tilde{\beta}(R)-\beta^0=(NT)^{-1/2}W_{NT}^{-1}\left(C_{NT}^{(1)}+C_{NT}^{(2)}\right)+\mathcal{R}_{NT}^R=\gamma_{NT}+\mathcal{R}_{NT}^R,$$

where the remainder term $\mathcal{R}_{NT}^R$ satisfies $\left\|\mathcal{R}_{NT}^R\right\|=O_P\left(\left\|\gamma_{NT}\right\|\delta_{NT}^{-1}\right)+o_P\left(\delta_{NT}^{-3}\right)=O_P\left(K^{1/2}T^{-1/2}\delta_{NT}^{-2}\right)+o_P\left(\delta_{NT}^{-3}\right)=o_P((NT)^{-1/2})$ under Assumptions A.3(i) and A.6(i).

Step 5. We show (v) $\sqrt{NT/K}\left\|\tilde{\beta}^c(R)\right\|=O_P(1)$ and $\sqrt{NT}\tilde{\beta}_k^c(R)=O_P(1)$ for each $k=1,2,...,K$. Comparing the results in Steps 2 and 4, we notice that $\tilde{\beta}(R)-\beta^0$ share the same asymptotic bias as $\tilde{\beta}(R_0)-\beta^0$. So the asymptotic analysis in MWb can be used to show that $\sqrt{NT/K}\left\|\tilde{\beta}^c(R)\right\|=O_P(1)$ and $\sqrt{NT}\tilde{\beta}_k^c(R)=O_P(1)$ for each $k=1,2,...,K$. The major difference is that MWb only consider fixed K but we allow slowly diverging K. Here, we outline the major steps only.

Recall $C_{NT}^{(2)}=C_{NT}^{(2,1)}+C_{NT}^{(2,2)}+C_{NT}^{(2,3)}$. We want to show to show that $C_{NT}^{(1)}$ contributes to both the asymptotic bias and variance of $\tilde{\beta}(R)$, $C_{NT}^{(2,1)}$ and $C_{NT}^{(2,2)}$ contribute to the asymptotic bias, and $C_{NT}^{(2,3)}$ is asymptotically negligible.

(i) First, we want to show that $C_{NT}^{(2,3)}$ is asymptotically negligible by showing that $\left\|C_{NT}^{(2,3)}\right\|=o_P(1)$. Using $\Phi=\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(F^{0\prime}F^0\right)^{-1}F^{0\prime}$, $M_{F^0}=I_T-P_{F^0}$, $|\mathrm{tr}(A)|\leq$rank$(A)\left\|A\right\|_{\mathrm{sp}}$, and CS inequality,

$$\begin{aligned}
\left\|C_{NT}^{(2,3)}\right\|^2 &= \sum_{k=1}^K\left[C_{NT,k}^{(2,3)}\right]^2=\frac{1}{NT}\sum_{k=1}^K\left[\mathrm{tr}\left(F^{0\prime}\boldsymbol{\varepsilon}'M_{\lambda^0}\mathbf{X}_kM_{F^0}\boldsymbol{\varepsilon}'\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(F^{0\prime}F^0\right)^{-1}\right)\right]^2 \\
&\leq \frac{2R_0}{NT}\sum_{k=1}^K\left\|F^{0\prime}\boldsymbol{\varepsilon}'M_{\lambda^0}\mathbf{X}_k\boldsymbol{\varepsilon}'\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(F^{0\prime}F^0\right)^{-1}\right\|_{\mathrm{sp}}^2 \\
&\quad +\frac{2R_0}{NT}\sum_{k=1}^K\left\|F^{0\prime}\boldsymbol{\varepsilon}'M_{\lambda^0}\mathbf{X}_kP_{F^0}\boldsymbol{\varepsilon}'\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(F^{0\prime}F^0\right)^{-1}\right\|_{\mathrm{sp}}^2\leq 2\mathcal{I}_1+2\mathcal{I}_2,\ \text{say.}
\end{aligned}$$

Using $M_{\lambda^0} = I_N - P_{\lambda^0}$, by Assumptions A.1(ii)-(iii), A.4(iii), S.1(i)-(ii), and A.6(i), we have

$$
\begin{aligned}
\mathcal{I}_1 &\leq \frac{R_0}{NT}\left\|\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(F^{0\prime}F^0\right)^{-1}\right\|^2\sum_{k=1}^{K}\left\|F^{0\prime}\boldsymbol{\varepsilon}' M_{\lambda^0}\mathbf{X}_k\boldsymbol{\varepsilon}'\lambda^0\right\|_{\mathrm{sp}}^2 \\
&\leq \frac{R_0}{NT}\left\|\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(F^{0\prime}F^0\right)^{-1}\right\|^2\frac{R_0}{(NT)^2}\sum_{k=1}^{K}\left\|F^{0\prime}\boldsymbol{\varepsilon}'\mathbf{X}_k\boldsymbol{\varepsilon}'\lambda^0\right\|_{\mathrm{sp}}^2 \\
&\quad +\frac{R_0}{NT}\left\|\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(F^{0\prime}F^0\right)^{-1}\right\|^2\sum_{k=1}^{K}\left\|F^{0\prime}\boldsymbol{\varepsilon}' P_{\lambda^0}\mathbf{X}_k\boldsymbol{\varepsilon}'\lambda^0\right\|_{\mathrm{sp}}^2 \\
&= \frac{R_0}{NT}\left\|\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(F^{0\prime}F^0\right)^{-1}\right\|^2\sum_{k=1}^{K}\left\|F^{0\prime}\boldsymbol{\varepsilon}'\mathbf{X}_k\boldsymbol{\varepsilon}'\lambda^0\right\|_{\mathrm{sp}}^2 \\
&\quad +\frac{R_0}{NT}\left\|\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(F^{0\prime}F^0\right)^{-1}\right\|^2\left\|\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\right\|^2\left\|F^{0\prime}\boldsymbol{\varepsilon}'\lambda^0\right\|_{\mathrm{sp}}^2\sum_{k=1}^{K}\left\|\mathbf{X}_k\boldsymbol{\varepsilon}'\lambda^0\right\|_{\mathrm{sp}}^2 \\
&= (NT)^{-1}O_P\left((NT)^{-2}\right)O_P\left(K(NT)^2(N+T)\right) \\
&\quad +(NT)^{-1}O_P\left((NT)^{-2}\right)O_P\left(N^{-1}\right)O_P\left(NT(N+T)\right)O_P\left(KNT(N+T)\right) \\
&= O_P\left(K\left(\delta_{NT}^{-2}+TN^{-2}\right)\right) = o_P(1).
\end{aligned}
$$

Similarly, we can readily show that $\mathcal{I}_2 = O_P\left(K\left(\delta_{NT}^{-2}+TN^{-2}\right)\right) = o_P(1)$. It follows that $\left\|C_{NT}^{(2,3)}\right\| = o_P(1)$.

(ii) Following Footnote 18 in the main text, we can write $C_{NT}^{(1)} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\tilde{X}_i' M_{F^0}\varepsilon_i$, where $\tilde{X}_i = X_i - \mathcal{X}_{i2NT}$ and $\mathcal{X}_{i2NT} = \frac{1}{N}\sum_{j=1}^{N}\lambda_i^{0\prime}\left(N^{-1}\lambda^{0\prime}\lambda^0\right)^{-1}\lambda_j^0 X_j$. We can make the following decomposition:

$$
\begin{aligned}
C_{NT}^{(1)} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left(X_i - M_{F^0}\mathcal{X}_{i2NT}\right)'\varepsilon_i - \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}X_i' P_{F^0}\varepsilon_i \\
&= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[X_i - P_{F^0}E_{\mathcal{D}}(X_i) - M_{F^0}\mathcal{X}_{i2NT}\right]'\varepsilon_i - \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[X_i - E_{\mathcal{D}}(X_i)\right]' P_{F^0}\varepsilon_i \\
&\equiv V_{NT} - B_{NT}^{(1)},\ \text{say}
\end{aligned}
$$

where the first term V_{NT} contributes to the asymptotic variance and the second term $B_{NT}^{(1)}$ contributes to the asymptotic bias. Note that $B_{NT}^{(1)}$ is defined analogously to $\mathbb{B}_{4NT}$ in Section 3.3 with X_i replaced by $X_{i(1)}$. Following Step 4 in the proof of Corollary 3.4, we can readily show that $\left\|\tilde{B}_{NT}^{(1)} - B_{NT}^{(1)}\right\| = o_P(1)$. In addition, we can show that $\|V_{NT}\| = O_P\left(K^{1/2}\right)$.

(iii) Using $M_{F^0} = I_T - P_{F^0}$, $\mathrm{tr}(A) = \mathrm{tr}(A')$, and $M_{\lambda^0} = I_N - P_{\lambda^0}$, we can make the following decomposition:

$$
\begin{aligned}
C_{NT,k}^{(2,1)} &= (NT)^{-1/2}\,\mathrm{tr}\left(\boldsymbol{\varepsilon}P_{F^0}\boldsymbol{\varepsilon}' M_{\lambda^0}\mathbf{X}_k\Phi'\right) - (NT)^{-1/2}\,\mathrm{tr}\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' M_{\lambda^0}\mathbf{X}_k\Phi'\right) \equiv C_{NT,k}^{(2,1)}(1) - C_{NT,k}^{(2,1)}(2), \\
C_{NT,k}^{(2,2)} &= -(NT)^{-1/2}\,\mathrm{tr}\left(\Phi'\mathbf{X}_k M_{F^0}\boldsymbol{\varepsilon}' M_{\lambda^0}\boldsymbol{\varepsilon}\right) = -(NT)^{-1/2}\,\mathrm{tr}\left(\boldsymbol{\varepsilon}' M_{\lambda^0}\boldsymbol{\varepsilon}M_{F^0}\mathbf{X}_k'\Phi\right) \\
&= (NT)^{-1/2}\,\mathrm{tr}\left(\boldsymbol{\varepsilon}' P_{\lambda^0}\boldsymbol{\varepsilon}M_{F^0}\mathbf{X}_k'\Phi\right) - (NT)^{-1/2}\,\mathrm{tr}\left(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}M_{F^0}\mathbf{X}_k'\Phi\right) \equiv C_{NT,k}^{(2,2)}(1) - C_{NT,k}^{(2,2)}(2).
\end{aligned}
$$

It follows that $C_{NT}^{(2,s)}=C_{NT}^{(2,s)}(1)-C_{NT,k}^{(2,s)}(2)$ for $s=1,2$, where $C_{NT}^{(2,s)}(l)$ is obtained by stacking $C_{NT,k}^{(2,s)}(l)$ into a $K\times 1$ vector for $l=1,2$ and $s=1,2$. As in (i), it is easy to show that $\left\|C_{NT}^{(2,s)}\right\|=o_P(1)$. $C_{NT}^{(2,s)}(2)$ needs to be corrected for $s=1,2$. Following the proof of Theorem 4.4 in MWb, we can also show that $\left\|\hat{B}_{NT}^{(2)}-C_{NT}^{(2,1)}(2)\right\|=o_P(1)$, and $\left\|\hat{B}_{NT}^{(3)}-C_{NT}^{(2,2)}(2)\right\|=o_P(1)$. Then we can

$$\left\|\hat{B}_{NT}^{(2)}+C_{NT}^{(2,1)}\right\|=o_P(1),\text{ and }\left\|\hat{B}_{NT}^{(3)}+C_{NT}^{(2,2)}\right\|=o_P(1).$$

(iv) As in the proof of Theorem 3.3, we have $\left\|\tilde{W}_{NT}-W_{NT}\right\|_{\text{sp}}\leq (NT)^{-1}\sum_{i=1}^N\left\|\tilde{X}_i\right\|^2\|P_{\tilde{F}}-P_{F^0}\|=O_P\left(K\delta_{NT}^{-1}\right)$ by using the fact that $\|P_{\tilde{F}}-P_{F^0}\|=O_P\left(\delta_{NT}^{-1}\right)$ and Assumption A.1(viii). But this bound is not tight. For the purpose of bias correction, we need to strengthen this result. In fact, we can show that $\left\|\tilde{W}_{NT}-W_{NT}\right\|_{\text{sp}}=O_P\left(\delta_{NT}^{-1}\right)$. To see this, we first observe that

$$\left\|\tilde{W}_{NT}-W_{NT}\right\|_{\text{sp}}=\left\|(NT)^{-1}\sum_{i=1}^N\tilde{X}_i'\left(P_{\tilde{F}}-P_{F^0}\right)\tilde{X}_i\right\|_{\text{sp}}=\max\left\{|\varrho_{1NT}|,\ |\varrho_{2NT}|\right\},$$

where $\varrho_{1NT}=\mu_{\max}\left((NT)^{-1}\sum_{i=1}^N\tilde{X}_i'\left(P_{\tilde{F}}-P_{F^0}\right)\tilde{X}_i\right)$, and $\varrho_{2NT}=\mu_{\min}\left((NT)^{-1}\sum_{i=1}^N\tilde{X}_i'\left(P_{\tilde{F}}-P_{F^0}\right)\tilde{X}_i\right)$.

Then noting that $\text{tr}(AB)\leq\|A\|\,\|B\|$, we have

$$\begin{aligned}
|\varrho_{1NT}| &= \left|\max_{\|\varkappa\|=1}\left((NT)^{-1}\varkappa'\sum_{i=1}^N\tilde{X}_i'\left(P_{\tilde{F}}-P_{F^0}\right)\tilde{X}_i\varkappa\right)\right|=\left|\max_{\|\varkappa\|=1}\text{tr}\left\{\left(P_{\tilde{F}}-P_{F^0}\right)(NT)^{-1}\sum_{i=1}^N\tilde{X}_i\varkappa\varkappa'\tilde{X}_i'\right\}\right| \\
&\leq \|P_{\tilde{F}}-P_{F^0}\|\max_{\|\varkappa\|=1}\left\|(NT)^{-1}\sum_{i=1}^N\tilde{X}_i\varkappa\varkappa'\tilde{X}_i'\right\|=O_P\left(\delta_{NT}^{-1}\right)O_P(1)=O_P\left(\delta_{NT}^{-1}\right)
\end{aligned}$$

because

$$\begin{aligned}
\max_{\|\varkappa\|=1}\left\|(NT)^{-1}\sum_{i=1}^N\tilde{X}_i\varkappa\varkappa'\tilde{X}_i'\right\|^2 &= \max_{\|\varkappa\|=1}(NT)^{-2}\,\text{tr}\left(\sum_{i=1}^N\tilde{X}_i\varkappa\varkappa'\tilde{X}_i'\sum_{j=1}^N\tilde{X}_j\varkappa\varkappa'\tilde{X}_j'\right) \\
&= \max_{\|\varkappa\|=1}(NT)^{-2}\,\text{tr}\left(\sum_{j=1}^N\varkappa'\tilde{X}_j'\sum_{i=1}^N\tilde{X}_i\varkappa\varkappa'\tilde{X}_i'\tilde{X}_j\varkappa\right) \\
&\leq \max_{\|\varkappa\|=1}(NT)^{-2}\left(\sum_{j=1}^N\varkappa'\tilde{X}_j'\sum_{i=1}^N\tilde{X}_i\tilde{X}_i'\tilde{X}_j\varkappa\right) \\
&\leq \mu_{\max}\left(\frac{1}{NT}\sum_{i=1}^N\tilde{X}_i\tilde{X}_i'\right)\max_{\|\varkappa\|=1}\frac{1}{NT}\sum_{j=1}^N\varkappa'\tilde{X}_j'\tilde{X}_j\varkappa \\
&= \left[\mu_{\max}\left(\frac{1}{NT}\sum_{i=1}^N\tilde{X}_i\tilde{X}_i'\right)\right]^2\leq\left[\mu_{\max}\left(\frac{1}{NT}\sum_{i=1}^N X_iX_i'\right)\right]^2 \\
&= O_P(1),
\end{aligned}$$

where the first and second inequalities follow from the fact $A'BA\leq\mu_{\max}(B)\,A'A\leq\|B\|\,A'A$, the third equality follows from the fact $\mu_{\max}(AA')=\mu_{\max}(A'A)$ and the last equality follows from Assumption A.1(viii). Analogously, $|\varrho_{2NT}|=O_P\left(\delta_{NT}^{-1}\right)$. It follows that $\left\|\tilde{W}_{NT}-W_{NT}\right\|_{\text{sp}}=O_P\left(\delta_{NT}^{-1}\right)$.

(v) Using $\tilde{\beta}^c(R) = \tilde{\beta}(R) + (NT)^{-1/2}\tilde{W}_{NT}^{-1}(\tilde{B}_{NT}^{(1)} + \tilde{B}_{NT}^{(2)} + \hat{B}_{NT}^{(3)})$ and $\tilde{\beta}(R) - \beta^0 = (NT)^{-1/2} W_{NT}^{-1}(C_{NT}^{(1)} + C_{NT}^{(2)}) + \mathcal{R}_{NT}^{R_0}$, we have

$$\begin{aligned}
\tilde{\beta}^c(R) - \beta^0 &= \tilde{\beta}(R) - \beta^0 + (NT)^{-1/2}\tilde{W}_{NT}^{-1}\left(\tilde{B}_{NT}^{(1)} + \tilde{B}_{NT}^{(2)} + \hat{B}_{NT}^{(3)}\right) \\
&= (NT)^{-1/2} W_{NT}^{-1}\left(V_{NT} - B_{NT}^{(1)} + C_{NT}^{(2,1)} + C_{NT}^{(2,2)} + C_{NT}^{(2,3)}\right) \\
&\quad + (NT)^{-1/2}\tilde{W}_{NT}^{-1}\left(\tilde{B}_{NT}^{(1)} + \tilde{B}_{NT}^{(2)} + \hat{B}_{NT}^{(3)}\right) + \mathcal{R}_{NT}^{R_0} \\
&= (NT)^{-1/2} W_{NT}^{-1} V_{NT} \\
&\quad + (NT)^{-1/2}\tilde{W}_{NT}^{-1}\left[\left(-B_{NT}^{(1)} + \tilde{B}_{NT}^{(1)}\right) + \left(C_{NT}^{(2,1)} + \tilde{B}_{NT}^{(1)}\right) + \left(C_{NT}^{(2,2)} + \tilde{B}_{NT}^{(1)}\right) + C_{NT}^{(2,3)}\right] \\
&\quad + (NT)^{-1/2}\left(\tilde{W}_{NT}^{-1} - W_{NT}^{-1}\right)\left(-B_{NT}^{(1)} + C_{NT}^{(2,1)} + C_{NT}^{(2,2)} + C_{NT}^{(2,3)}\right) + \mathcal{R}_{NT}^{R_0} \\
&\equiv \mathcal{I}_3 + \mathcal{I}_4 + \mathcal{I}_5 + \mathcal{R}_{NT}^{R_0}, \text{ say.}
\end{aligned}$$

We can readily show that $\|\mathcal{I}_3\| = O_P\left((NT/K)^{-1/2}\right)$, $\|\mathcal{I}_4\| = o_P\left((NT)^{-1/2}\right)$, and $\left\|\mathcal{R}_{NT}^{R_0}\right\| = o_P\left((NT)^{-1/2}\right)$ by (i)-(iii) and Step 4. Let c_K be an arbitrary $K \times 1$ vector with $\|c_K\| = 1$. Then

$$\begin{aligned}
\left|c_K' B_{NT}^{(1)}\right| &= \frac{1}{\sqrt{NT}}\left|\operatorname{tr}\left(\sum_{i=1}^N \varepsilon_i c_K' [X_i - E_{\mathcal{D}}(X_i)]' P_{F^0}\right)\right| \\
&= \frac{1}{\sqrt{NT}}\left|\operatorname{tr}\left(\sum_{i=1}^N F^{0\prime}\varepsilon_i c_K' [X_i - E_{\mathcal{D}}(X_i)]' F^0 \left(F^{0\prime}F^0\right)^{-1}\right)\right| \\
&\le \frac{R_0}{\sqrt{NT}}\left\|\left(F^{0\prime}F^0\right)^{-1}\right\|\left\|\sum_{i=1}^N F^{0\prime}\varepsilon_i c_K' [X_i - E_{\mathcal{D}}(X_i)]' F^0\right\| \\
&\le \frac{R_0}{\sqrt{NT}}\left\|\left(F^{0\prime}F^0\right)^{-1}\right\|\left\{\sum_{i=1}^N \left\|F^{0\prime}\varepsilon_i\right\|^2 \sum_{i=1}^N \left\|[X_i - E_{\mathcal{D}}(X_i)]' F^0\right\|^2\right\}^{1/2} \\
&\le O_P\left(N^{-1/2}T^{-3/2}\right) O_P\left(K^{1/2}NT\right) = O_P\left(K^{1/2}N^{1/2}T^{-1/2}\right).
\end{aligned}$$

It follows that $(NT)^{-1/2}\left\|B_{NT}^{(1)}\right\| = (NT)^{-1/2} O_P\left(K^{1/2}N^{1/2}T^{-1/2}\right) = O_P\left(K^{1/2}T^{-1}\right)$. Similarly, we can show that $(NT)^{-1/2}\left\|C_{NT}^{(2,1)}\right\| = O_P\left(K^{1/2}N^{-1}\right)$ and $(NT)^{-1/2}\left\|C_{NT}^{(2,2)}\right\| = O_P\left(K^{1/2}T^{-1}\right)$. These results, in conjunction with the analysis in (i) and Assumption A.6(i), imply that

$$\|\mathcal{I}_5\| = O_P\left(\delta_{NT}^{-1}\right) O_P\left(K^{1/2}\delta_{NT}^{-2} + o_P(NT)^{-1/2}\right) = O_P\left(K^{1/2}\delta_{NT}^{-3}\right) = o_P\left((NT)^{-1/2}\right).$$

It follows that

$$\sqrt{NT}(\tilde{\beta}^c - \beta^0) = W_{NT}^{-1}V_{NT} + \mathbf{o}_P(1) \text{ and } \left\|\tilde{\beta}^c(R) - \beta^0\right\| = O_P((NT/K)^{-1/2}).$$

Lastly, let ι_k be the kth column of the identity matrix I_K. Then

$$\begin{aligned}
\left|\tilde{\beta}_k^c(R) - \beta_k^0\right| &= \left|\iota_k'\left(\tilde{\beta}^c(R) - \beta^0\right)\right| = \iota_k'\mathcal{I}_3 + o_P\left((NT)^{-1/2}\right) \\
&= O_P\left((NT)^{-1/2}\right) + o_P\left((NT)^{-1/2}\right) = O_P\left((NT)^{-1/2}\right)
\end{aligned}$$

because $(NT)^{1/2}\,\iota_k'\mathcal{I}_3 = \iota_k' W_{NT}^{-1} V_{NT} = \iota_k' W_0^{-1} V_{NT}\,\{1 + o_P(1)\} = O_P(1)$ by second moment calculations and Chebyshev inequality.

REFERENCES

Andrews, D. W. K., 2005. Cross-section regression with common shocks. Econometrica 73, 1551-1585.

Bai, J., 2003. Inferential theory for factors models of large dimensions. Econometrica 71, 135-173.

Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica 77, 1229-1279.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191-221.

Hahn, J., Kuersteiner, G., 2011. Reduction for dynamic nonlinear panel models with fixed effects. Econometric Theory 27, 1152-1191.

Moon, H., Weidner, M., 2014a. Linear regression for panel with unknown number of factors as interactive fixed effects. Working paper, University of Southern California.

Moon, H., Weidner, M., 2014b. Dynamic linear panel regression models with interactive fixed effects. Working paper, University of Southern California.

Pesaran, M. H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74, 967-1012.

Pollard, D., 1984. Convergence of Stochastic Processes. Springer-Verlag, New York.

Prakasa Rao, B. L. S., 2009. Conditional independence, conditional mixing and conditional association. Annals of the Institute of Statistical Mathematics 61, 441-460.

Roussas, G. G., 2008. On conditional independence, mixing and association. Stochastic Analysis and Applications 26, 1274-1309.

Stewart, G. W., and Sun, J.-G., 1990. Matrix Perturbation Theory. Academic Press, Boston.

Su, L., Chen, Q., 2013. Testing homogeneity in panel data models with interactive fixed effects. Econometric Theory 29, 1079-1135.

Su, L., Zhang, Y., 2014. Nonparametric dynamic panel data models with interactive fixed effects: sieve estimation and specification testing. Working paper, Singapore Management University.